Engineering Intracellular Antibody Libraries

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Chemical Engineering

University of Saskatchewan

Saskatoon

By

Wendy Lynn Bernhard

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Chemical Engineering

University of Saskatchewan

Saskatoon, Saskatchewan

Abstract

The goal of this research is to understand how three different parameters affect single chain variable fragment (scFv) binding capacity. The parameters that were varied include the number of variable complementarity determining regions (CDRs), the number amino acids used to diversify CDRs, and configuration of the structure. How the parameters affect the binding capacity will be tested using the yeast two hybrid assay against five different protein domains. Eight scFv libraries were generated; the genes expressing the scFvs were constructed and the CDRs were randomized using PCR amplification. Genes expressing scFvs were cloned, using the homologous gap repair mechanism in *Saccharomyces cerevisiae*. Representative members of scFv libraries were sequenced to confirm correct construction.

Library diversity was calculated from the library transformation efficiency. Transformation efficiency refers to the number of cells that grew at the time of transformation of the scFv gene into yeast cells. There were significant differences in the diversity of the scFv libraries, which created difficulty in comparing the library binding capacities. Sequencing the scFv libraries revealed that on average 50% of each library contained correct scFv sequences. The percent of correct sequences within each library was then used to calculate the functional diversity.

The yeast two-hybrid assay was used to screen the scFv libraries for interactions and to test binding capacity. The binding capacity of the scFv libraries was tested and compared in five different yeast two-hybrid assays using five protein domains as the targets for each screen.

The screening results showed that in all cases cyclic scFv libraries had a statistically significant higher binding capacity than linear scFv libraries despite a diversity bias against the cyclic libraries. There was no clear trend in binding capacity with the other two parameters; however, the four amino acid three CDR libraries dominated over the other libraries in almost every screen.

Some of the scFvs isolated from the screens were expressed in *E. coli* and *S. cerevisiae* to analyze for proper expression and correct size. All the scFvs that were isolated and analyzed were the correct size and could be purified using a poly histidine tag.

Due to its bioaffinity and specificity, scFvs were constructed to profile disease patterns, and to identify potential drug targets. In addition to its original application to health-related studies, scFvs could also be extended to locate potential metabolic bottlenecks, to alter metabolic flux to enhance productivity, and regulate metabolic bionetworks. Industrial microorganisms are generally carrying more than two sets of chromosomes, making it difficult to be genetically engineered when conventional approaches are employed. With the availability of scFvs as reported in this thesis, we are able to design specific scFvs that selectively bind to target proteins, resulting in re-routing of metabolic flux within the microorganism, toward a high productivity of desired product. ScFvs can be applied to industrial microorganisms directly, leading to the development of new fermentation processes.

Acknowledgements

I would like to sincerely thank my supervisor Dr. Yen-Han Lin for all his support and guidance through my studies and for supporting this collaboration. I am very grateful and fortunate to have received his continuous encouragement and feedback.

I would like to thank my co-supervisors Drs. C. Ron Geyer and Steve Sanche for their continuous support. Dr. Geyer initially hired me as a summer student; his novel ideas and enthusiasm encouraged me to pursue my Masters degree. His mentorship and direction has played a major role in my training and success through my graduate program.

My graduate committee including Drs Catherine Niu and Mehdi Nemati for their continuous support and feedback on my research project and reports.

I would also like to thank all the members of Dr. Geyer and Dr. DeCoteau's lab, especially Kris Barreto who helped with numerous aspects of my project; Dr. John DeCoteau for allowing me to use equipment and reagents from his lab and for his insightful comments during lab meetings. Karen Mochoruk for assistance with the mass spectrometer and making sure I always had reagents to do my experiments.

Lastly, I would like to send a special thanks to my family who have always supported and encouraged me along the way.

Table of Contents

page

List of Tables

List of Figures

List of Abbreviations

Ade2	Gene involved in adenine biosynthesis
BCR-ABL	Breakpoint cluster region – ableson kinase
CC1-72	Coiled–coil domain, amino acids 1-72 of BCR
CDR	Complementarity determining region
cK3	Cyclic, four amino acid three CDR scFv library
cK4	Cyclic, four amino acid four CDR scFv library
cT3	Cyclic, two amino acid three CDR scFv library
cT4	Cyclic, two amino acid four CDR scFv library
dNTPs	DeoxyNucleotide TriPhosphates
Fab	Fragment antigen binding
H1	Complementarity determining region one on the heavy chain
H2	Complementarity determining region two on the heavy chain
H3	Complementarity determining region three on the heavy chain
His3	Gene involved in histidine biosynthesis
I_C	C-terminal domain of the intein
I_N	N-terminal domain of the intein
L1	Complementarity determining region one on the light chain
L2	Complementarity determining region two on the light chain
L3	Complementarity determining region three on the light chain

lacZ	Reporter gene encoding ß-galactosidase
Leu2	Gene involved in leucine biosynthesis
LK3	Linear four amino acid three CDR scFv library
LK4	Linear four amino acid four CDR scFv library
LT3	Linear two amino acid three CDR scFv library
LT4	Linear two amino acid four CDR scFv library
MATa	*S. cerevisiae* mating type a
MATα	*S. cerevisiae* mating type α
Npu	*Nostoc punctiforme*
Oligo	Oligonucleotide
PCR	Polymerase chain reaction
ScFv	Single chain variable fragment
SD	Synthetic dropout
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH2	Src homology 2
SH3	Src homology 3
Ssp	*Synechocystic species*
Taq	*Thermos aquaticus*
Trp1	Gene involved in tryptophan biosynthesis
Ura2	Gene involved in uracil biosynthesis
Ura3	Gene involved in uracil biosynthesis
YPD(A+)	Yeast peptone dextrose (with adenine added)

Glossary

::	means “replaced by” in describing strains, where lowercase lettering indicates a gene that is knocked out, uppercase indicates a functional gene; For example: “leu2::LexA6op-LEU2” means leu2 was replaced by six LexA operons and the Leu2 gene
Ade2	Gene encoding the enzyme phosphoribosylaminoimidazole carboxylase which catalyzes a step in the purine nucleotide biosynthetic pathway
Ampicillin	Antibiotic used in plasmid selection in *E. coli*.
Anti-β-galactosidase	Refers to an antibody or scFv that is specific for β-galactosidase
Antibody	Protein produced by the immune system; removes foreign organisms and objects by binding them with high specificity and affinity.
Antigen	Antibody or antibody fragment target
Bait	In a yeast two-hybrid screen, the target protein or protein of interest screened against other proteins to find an interacting protein.
BCR-ABL	A tyrosine-kinase oncogene expressed in chronic myeloid leukaemia; created from the Philadelphia-chromosome translocation; the Abelson leukaemia-virus protein (ABL) is fused with the breakpoint-cluster region (BCR).
CC1-72	Coiled coil domain, BCR-ABL protein domain that forms a bundle of two or three alpha helices and can be involved in protein interactions
CDR	Complementarity Determining Region, found on antibodies and scFvs, forms loops that have affinity and specificity for the target antigen
cK3	Cyclic scFv library containing three variable heavy chain CDRs and one variable light chain CDR (L3) varied with two different amino acids, tyrosine, serine, aspartate and alanine.

cK4	Linear scFv library containing three variable heavy chain CDRs and one variable light chain CDR (L3) varied with four different amino acids, tyrosine, serine, aspartate and alanine.
Clone	To insert a gene of interest into a vector for expression inside a cell.
Codon	Three letter code made from four different DNA bases which code for different amino acids.
Co-transformation	Introduction of more than one foreign DNA molecules into a cell.
cT3	Cyclic scFv library containing three variable heavy chain CDRs varied with two different amino acids, tyrosine and serine.
cT4	Cyclic scFv library containing three variable heavy chain CDRs and one variable light chain CDR (L3) varied with two different amino acids, tyrosine and serine.
Degenerate	In reference to codons, a codon that can code for more than one amino acid.
Diploid	A yeast cell consisting of both mating types a and α; will also express two different plasmids, that of an scFv and that of a target protein domain.
dNTPs	DeoxyNucleotide TriPhosphates or DNA monomers.
*Eco*RI	Restriction enzyme that cuts the DNA sequence G|AATTC between the G and A, isolated from *E. coli* RY13.
Electrocompetent	A bacterial cell that can be transformed using electroporation.
Electrophoresis	Migration of a DNA sample through a medium under the influence of an electric field to separate the different DNA samples.
Electroporation	A way to transform bacterial cells using an electrical shock that induces t transient pores in the cells through which DNA can enter.
Endonulcease	Enzyme that specifically cleaves the phosphodiester bond in DNA.
Fab	Fragment antigen binding; an antibody fragment composed of the N-terminal half of the heavy chain and its associated light chain.
Functional scFv	An scFv containing the proper sequences determined by sequencing.
Gel purification	To separate DNA through an agarose gel, cut the desired DNA out of the gel and clean it up to obtain pure DNA from other contaminating DNA.

H1/H2/H3	Refers to a specific CDR region within the heavy chain (ie H2 refers to CDR 2 on the heavy chain).
Hck	Hemopoietic cell kinase, may be involved in chronic myeloid leukaemia.
Heavy chain	The large polypeptide subunit of an antibody.
His3	Gene encoding imidazole glycerol-phosphate dehydratase, which catalyzes the sixth step in histidine biosynthesis.
Histidine tag	Five or six consecutive histidines attached to a protein; has high affinity to nickel which is used to purify the tagged protein.
Homologous Recombination	A natural DNA gap repair system found in yeast; can be used to clone
I_C Domain	The I_C domain of the Ssp self splicing split intein.
I_N Domain	The I_N domain of the Npu self splicing split intein.
Intein	In this study the Ssp I_C domain and Npu I_N domain were used to catalyze a cyclization reaction cyclizing the scFv.
Kanamycin	Antibiotic used in plasmid selection in *E. coli*.
L1/L2/L3	Refers to a specific CDR region within the light chain (ie H2 refers to CDR 2 on the light chain).
lacZ	Reporter gene that encodes β-galactosidase.
Leu2	LEU2 encodes the enzyme beta-isopropylmalate dehydrogenase, the en catalyzes the third step in leucine biosynthesis.
LexA operator	DNA sequence that LexA binds to with high efficiency.
LexA8op	Containing eight LexA operators.
Library	(in this study) A collection of yeast cells each expressing an scFv with a different CDR sequence
Ligase	An enzyme that joins two DNA molecules together.
Ligation	The joining of two DNA molecules using a ligase.
Light chain	The small polypeptide subunit of an antibody.

Linker	Short peptide made up of a three [Ser(Gly4)]3 used to tether the heavy chain to the light chain and to aid in cyclization of scFv proteins by allowing scFv conformation to stay unchanged.
LK3	Linear scFv library containing three variable heavy chain CDRs varied with four different amino acids, tyrosine, serine, aspartate and alanine.
LK4	Linear scFv library containing three variable heavy chain CDRs and one variable light chain CDR (L3) varied with four different amino acids, tyrosine, serine, aspartate and alanine.
LT3	Linear scFv library containing three variable heavy chain CDRs varied with two different amino acids, tyrosine and serine.
LT4	Linear scFv library containing three variable heavy chain CDRs and one variable light chain CDR (L3) varied with two different amino acids, tyrosine and serine.
Lysate	All material left from the lysis of a cell.
MATa	*S. cerevisiae* mating type a.
MATα	*S. cerevisiae* mating type α
Npu	I_N domain from *Npu Dna*E (*Nostoc punctiforme,* strain PCC73102, DNA polymerase III alpha subunit).
Oligonucleotide (oligo or primer)	A short DNA sequence.
PCR	Polymerase Chain Reaction, the process of amplifying DNA
pESC-Hck	A plasmid expressing Hck.
Phage	A bacterial virus.
Phage display	A technique used to fuse and display a foreign peptide or protein on a phage cell surface. The phage are screened to see what targets bind to the expressed fusion protein displayed on the phage surface.
pIL500	Vector used to express scFv library proteins; Contains kanamycin resistance, and yeast two-hybrid machinery including reporter genes.
Protein L	A protein that binds to antibodies and scFvs through their light chain; isolated from the bacteria *Peptostreptococcus magnus*.

Recombination	way to introduce foreign DNA into a vector where two linear strands of DNA containing complementary sequences will join together to create one DNA molecule.
Reporter gene	Expression determines if an scFv interacts with a protein domain in the yeast two-hybrid system.
Restriction site	Specific sequence of DNA that is recognized and cut by a restriction enzyme in a specific location.
ScFv	Single chain variable fragment, intrabody or intracellular antibody; an antibody fragment that is made from a light chain variable domain and a heavy chain variable domain from an antibody.
ScFv library	Collection of yeast cell expressing scFv proteins varied by their CDR composition.
SD	Synthetic dropout media.
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis; used to separate proteins based on size.
SH2	Src homology 2; BCR-ABL protein domain that binds phosphotyrosine-containing polypeptide.
SH3	Src homology 3; BCR-ABL protein domain that binds to target proteins proline rich sequences with PXXP as the binding motif.
Ssp	Ic domain from Ssp DnaE (*Synechocystis species*, strain PCC6803, DnaE polymerase III alpha subunit).
Taq Polymerase	Synthesizes a complementary DNA strand from a single strand of DNA, isolated from *Thermos aquaticus*.
Thermocylcer	PCR machine or DNA amplifier, holds at different temperatures to allow DNA to denature, anneal, elongate.
Transformation	Introduction of foreign DNA into a cell.
Trp1	Gene encoding phosphoribosylanthranilate isomerase, an enzyme that catalyzes the third step in tryptophan biosynthesis.
Tyrosine Kinase	An enzyme that phosphorylates tyrosine molecules in proteins.

Ura2	Gene encoding the enzyme carbamoylphosphate synthetase - aspartate transcarbamylase which catalyze the first two enzymatic steps in the biosynthesis of pyrimidines.
Ura3	Gene encoding the enzyme orotidine 5-phosphate decarboxylase, which is involved in the synthesis of pyrimidine ribonucleotides.
Vector (or Plasmid)	Used to express foreign proteins or proteins as a higher concentration than normal in genetically modified cells.
Western blot	A nitrocellulose membrane that contains protein transferred from an electrophoretic gel; the membrane (or blot) is probed with a labeled antibody that will bind specifically to the protein of interest.
X-gal	Used to indicate whether a bacterium expresses the β-galactosidase enzyme.
*Xho*I	Restriction enzyme that cuts the DNA sequence C|TCGAG between the first C and T, isolated from *Xanthomonas campestris*.
Yeast two-hybrid	An assay used to identify protein-protein interactions within a yeast cell.
YPD	Yeast peptone dextrose; media for growing yeast.
YPDA+	Yeast peptone dextrose with adenine added; media for growing yeast.
β-galactosidase	Enzyme that hydrolyzes X-gal into galactose and 4-chloro-3-brom-indigo which forms a visible blue precipitate.

Chapter 1 Introduction

1.1 Background

Antibodies are proteins produced by the immune system that bind with high specificity and affinity to foreign molecules. The target of an antibody is called an antigen. The ability of antibodies to specifically bind to their target with high affinity makes them desirable as therapeutic, diagnostic, and research affinity reagents (1). Therefore, the production of functional antibodies and techniques to isolate them against protein targets is in high demand.

Typically, antibodies are produced by challenging a host animal with an antigen (2) or by screening synthetic antibody libraries for antibodies that bind a particular antigen (3). Raising antibodies in animals is time consuming and expensive. Single chain variable fragments (scFvs) can be screened for interactions using *in vitro* display systems such as phage, yeast surface, and ribosome display (4). The most common way to isolate synthetic antibodies is with phage display technology (3). Currently, it is not possible to express antibodies on the surface of phage due to their large size and folding limitations. As a result, synthetic antibody fragments containing parts of the antibody are used for phage display. One type of antibody fragment, referred to as single-chain variable fragments (scFvs), consists of the light and heavy variable domains of an antibody connected by a peptide linker; forming a single amino acid chain that ideally folds to reconstitute the binding site of a natural antibody. Despite lacking antibody constant regions, scFvs have extremely high affinity and specificity for their targets. This binding ability comes from the complementarity determining regions (CDRs). Each variable

domain contains three CDRs that fold to form the binding site of the scFv. CDRs are surrounded by constant regions that serve as a scaffold to correctly display them.

Phage display and other display technologies have proven to be useful for isolating scFvs that interact with extracellular targets. ScFvs isolated against intracellular targets often do not express well inside of cells, or bind poorly *in vivo*. Poor expression may be due to the structure of the scFvs, which consist primarily of a β-barrel framework formed by the hydrophobic interaction of the two variable domains (5). The interaction between the two variable domains is not inherently stable, as scFvs have been shown to unfold and aggregate. The interaction between these two domains is enhanced by the linker peptide, which keeps the domains in close proximity, but aggregation is still a problem. In the cell, this instability leads to low expression levels and short half lives of the scFvs.

Despite the relative instability of scFvs inside cells, there are several reports using the yeast two-hybrid assay as a selection assay to isolate scFvs against intracellular protein targets (6). The yeast two-hybrid assay is a common technique used to identify protein interactions of interest inside an eukaryotic cell. This serves as a technique to isolate scFvs under selective pressure to fold and function inside mammalian cells (7). In the yeast two-hybrid assay, two hybrid proteins are co-expressed from plasmids in *S. cerevisiae*. One hybrid protein consists of the target protein fused to a DNA binding domain, which binds to the promoter region of a reporter gene. The second hybrid protein contains a member of a protein library, in our case an scFv, fused to a transcription activation domain. If the target protein and a library member interact with each other, the transcription activation domain is brought to the promoter resulting in the activation of the reporter gene. The activity of the reporter gene(s) is monitored to select for protein interactions.

To bind any given protein target with high affinity, the scFv library must contain many members. The variable CDRs within the scFv library dictates the specificity of the individual scFvs. Mutating the amino acids in the CDRs results in combinatorial libraries of scFvs. On average there are approximately 20 positions within the CDRs that contact the target protein. Since there are 20 naturally occurring amino acids, a library containing every combination would have 20^{20} different members using all amino acids at each important positions within the CDRs. A library this large is impossible to create or screen in yeast. To overcome this problem, antibody fragment libraries can be created with decreased diversity by limiting the number of amino acids used at each position, or the number of amino acid positions mutated (8,9). To decrease the diversity of an antibody fragment library the amino acids varied within the CDRs (number of amino acids) and/or the number of variable CDRs (number of amino acid positions) are reduced (8,9). ScFv libraries with minimal diversity still contain high affinity binders that can be isolated against a range of targets. Library design is a crucial aspect in producing combinatorial peptide libraries because not all amino acids are specific in their interactions. Arginine in particular is often involved in non-specific binding (10). Where tyrosine and serine are usually involved in specific interactions (10) and tyrosine is primarily involved in contacting the antigen (8).

Another critical point in creating scFv libraries is the stability of scFvs. The scFv structure and systems to express scFvs have been modified in an attempt to increase the ability of scFvs to fold properly within the cell. Many of these modifications modestly improve scFv folding; however, an optimal system for expressing scFvs in cells has not yet been developed. For example, fusing scFvs to stable, soluble proteins can create steric hindrance, which prevents proper folding of the protein (11). Cyclization of scFvs has not been reported, but generally

protein cyclization has been shown to increase protein stability and solubility in a number of different studies (12-15). Cyclization constrains the protein in a conformation where the protein cannot unfold and aggregate. Cyclizing scFvs should be useful in increasing their stability ultimately increasing their ability to remain soluble.

1.2 Objectives

The objectives of this research were to optimize scFv library design by increasing the stability and affinity of scFvs isolated using the yeast two-hybrid assay. To achieve this objective the following three variables were characterized:

(i) CDR amino acid diversity;

(ii) Number of CDRs containing random amino acids; and

(iii) Cyclization of scFvs.

1.3 Approach

To investigate the effect of the three different parameters on the scFv folding and binding capacity in the yeast two-hybrid assay, eight different libraries were designed and generated. Six scFv libraries were initially produced LK3, LK4, cK4, LT3, LT4, cT4, where c is cyclic, L is linear, K is four amino acids, T is two amino acids, and the number following K or T is the number of CDRs. The six libraries were screened five times in five different yeast two-hybrid screens. In every screen, more scFvs were isolated from the cyclic libraries compared to the linear libraries. To confirm whether this trend is also observed with the K3 and T3 libraries, cK3 and cT3 libraries were subsequently synthesized. Individual scFvs isolated from the yeast two-hybrid screens were expressed in *S. cerevisiae* and *E. coli* to analyze the expression and to purify them for further analysis of cyclization.

1.4 Thesis Organization

In the first section of Chapter 2, natural antibodies and scFvs are introduced and scFv applications are reviewed. In the second section, the three variable parameters tested in this study are described in detail. In the third section, the yeast two-hybrid assay is explained as well as the protein domains used in the assay. Chapter 3 reviews the materials and methods use in this study. Chapter 4 presents and describes the results found from the scFv yeast two-hybrid assays, and scFv protein expression and purification. In Chapter 5, results obtained from the experiments are described. Chapter 6 the results are discussed with respect to the current literature and Chapter 7 contains future recommendations for this project.

Chapter 2 Literature Review

2.1 Natural antibodies and scFvs

Natural antibodies are large proteins that possess many desirable properties that can be exploited in medicine and research. They are made up of two identical heavy chains and two identical light chains (Figure 2-1). Each light chain contains one variable domain and one constant domain. Each heavy chain contains one variable domain and on average three constant domains. The constant domains hold the antibody together by disulphide bonds. The variable domains contain three hypervariable regions called, complementarity determining regions (CDRs). The CDRs fold up to form loops, which bind to the antibody's target and dictate specificity and affinity.

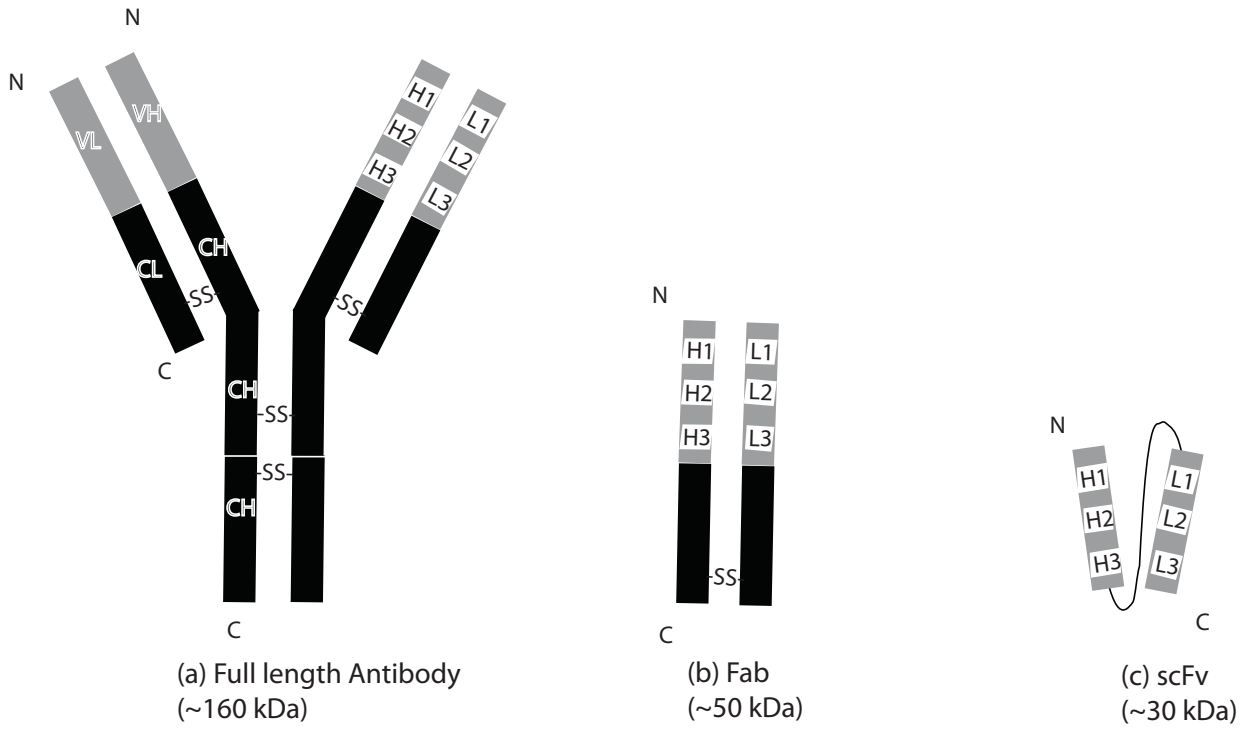


Figure 2-1: Full length antibody, Fab fragment and Fv fragments.
(a) The full-length antibody contains two identical light chains and two identical heavy chains. The antibody contains two variable light chain domains (VL), two constant light chain domains (CL), two variable heavy chain domains (VH) and three or four constant heavy chain domains (CH). The constant domains are held together and stabilized by disulphide bonds (-SS-). Each variable region contains three hypervariable regions called complementarity determining regions (CDRs). There are three CDRs on each heavy chain (H1, H3, H3) and three on each light chain (L1, L2, L3). (b) Fragment antigen binding (Fab) portion of an antibody. Fabs are made up of a VL, VH, CL and CH. (c) Single chain variable fragments (scFv) contain the CDRs of an antibody and are the smallest antibody fragment that retains the antibody-binding site. Figure adapted from (16).

Early studies have shown that antibodies can be used to block the function of their targets by injecting them into a population of cells (17). Antibodies cannot be readily used to block intracellular targets because their structure is stabilized by disulphide bonds, which do not form within the reducing environment of a cell (18). As a result, it is not possible to express properly folded, stable antibodies within cells. Also, the large size of the antibody structure limits its

intracellular use and has led to the development of smaller antibody fragments for intracellular applications (4).

Antibody fragments are created using the hypervariable binding regions of the antibody (Figure 2-1). The single-chain variable fragment (scFv) is one of the most commonly studied antibody fragments (Figure 2-2).

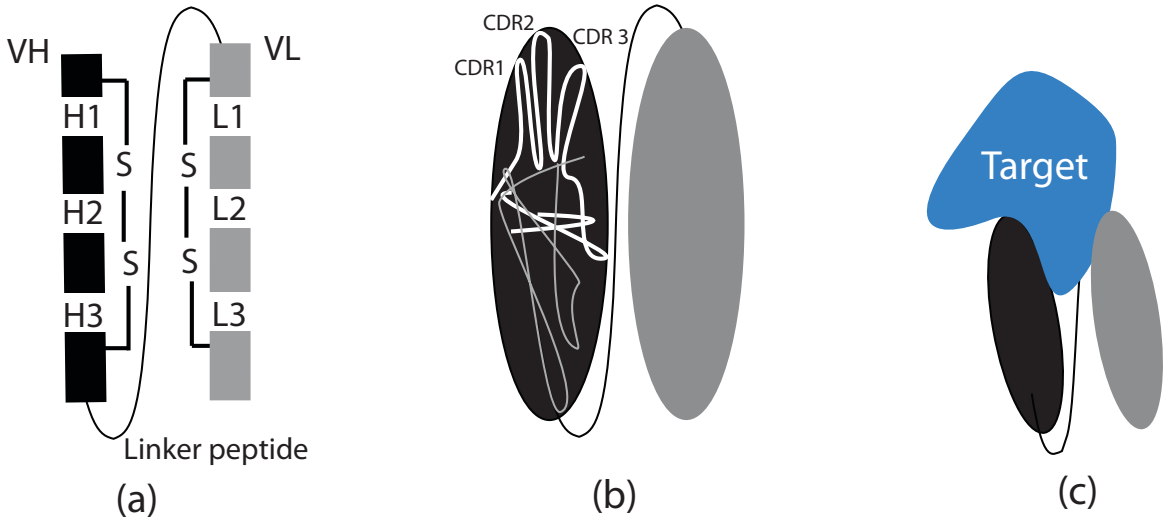


Figure 2-2**:** ScFv structure and CDR folding and binding model**.**
(a) Single chain variable fragment (scFv) consists of a variable light chain domain (VL) and a variable heavy chain domain (VH). The domains are linked together by a linker peptide. Each domain contains three complementarity determining regions (CDR). The domains are stabilized by inter-domain disulphide bonds (-S-S-). (b) The CDRs of the scFv fold forming loops which create the binding pocket. (c) An example of how a scFv binds to its antigen through its CDRs. Figure adapted from (16).

ScFvs, which are also referred to as intracellular antibodies or intrabodies, are a smaller, more versatile alternative to antibodies. ScFvs are made up of a heavy chain variable region and a light chain variable region of the natural antibody linked together by a linker peptide (Figure 2-

2) (19). The linker peptide constrains the two variable domains within the vicinity of each other, which increases the effective concentration of the light and heavy chains and stabilizing the scFv structure (20). The flexibility of the linker favors the intra-molecular association of the two variable domains without obstructing their interaction with each other. The variable domains of the heavy and light chains are further stabilized by disulphide bonds within the variable domains. Each variable fragment contains three CDRs, which fold up to form loops that contact the target protein and dictate the binding specificity (19).

2.2 Applications of scFvs

The development of scFvs has increased the applications for antibody-like reagents. The intracellular use of scFvs increases their application potential compared to the full-length antibody. ScFvs with identical CDRs to a natural antibody bind the same targets with similar binding affinity (19). This means that any full-length antibody isolated against a target can be made more versatile by creating its scFv counterpart. This versatility expands to fields such as therapeutics, diagnostics and research.

The use of scFvs in research has increased in the last few years. ScFvs are used in bio-affinity assays to specifically purify a target protein (21). ScFv microarrays can be used to identify biomarkers associated with specific diseases (22). Biomarkers can be used to understand, identify and diagnosis disease. They can also be used as drug targets. By identifying potential drug targets further studies can be done to develop drugs. For example, Carlsson *et al* use scFv microarrays to create a signature protein profile for metastatic breast cancer (22). These types of profiles allow the identification of proteins that may be mis-regulated in a malignancy. The mis-regulated proteins can then be used as targets in future studies to determine their role in

the disease. Proteins linked to the cause of a disease can be screened against libraries of scFvs. ScFvs isolated against disease-associated targets can be used to cure or treat the disease. ScFvs with high affinity and specificity to a target protein can be used to inhibit function by disrupting protein-substrate, protein-protein, DNA-protein interactions, localizing the target away from site of action, or targeting the protein or cell to death.

ScFvs can be used as diagnostic tools to identify potential drug targets. Protein microarrays can be used to identify protein-protein interactions, substrates of an enzyme, targets of biologically active small molecules in search of specific target proteins found in particular diseases (23). Protein microarrays made with scFvs are very useful because they are a high-throughput assay that can detect very low amounts of specific proteins within a mixture that may not be detectable by other methods due to low concentration. Early detection of specific biomolecules involved in certain diseases is important because it can make the difference between being able to treat a disease and the spread of the disease into an untreatable state. For example, a fungal infection in an immuno-compromised person can be fatal, however if the fungi is detected early enough when the fungus has not had as much time to multiply it can be more successfully treated. Improvements in detection and diagnostics are also useful future predictors of relapses and to monitor treatments (23). The output signal of protein arrays can be colorimetric, fluorescence, or chemiluminescence. Colorimetric assays developed using scFvs contain the specific scFv tethered to a protein tracer. For example, cells being tested for rabies viral infection can be immobilized and incubated with a rabies specific scFv fused to alkaline phosphatase (24). The excess scFv-alkaline phosphatase fusion is washed away and an alkaline phosphatase substrate, such as p-nitrophenol phosphate can be added. If the rabies virus is present in the cells they will bind to the scFv-alkaline phosphatase fusion. The alkaline

phosphatase will cleave the p-nitrophenol phosphate into p-nitrophenol, which can be monitored at 405 nm (24). This is a direct way to detect the presence of rabies virus. Fluorophores fused to scFvs are used in a similar way (25). Chemiluminescence could be used in protein arrays; however, to date no studies have been reported. These diagnostic measures can be used in the diagnosis of many diseases. They are inexpensive, rapid and very sensitive which is useful in early detection of a disease. With early detection of a disease it is easier to treat.

ScFvs have been isolated against various proteins to use in the treatment of disease. Their ability to penetrate tissue and bind their target is better because they are smaller than antibodies (26). ScFvs also clear faster from the body than larger antibody fragments making them less immunogenic (26). ScFvs can be used to target specific proteins or protein domains to inhibit or block protein function. Viral and host cell proteins causing infectious disease such as HIV-1 have been targeted with scFvs to try and control infection, replication and spread of the HIV-1 virus (reviewed in, (16)). ScFvs specific for oncogenes, fusion protein products and over-expressed regulatory proteins and receptors involved in the development and spread of cancer have been isolated (reviewed in, (16)). Neurological diseases caused by aggregated and hyper-phosphorylated proteins such as Huntington's disease and Alzheimer's have also been targeted by scFvs (reviewed in, (16)). Anti-sperm scFvs have been isolated with possible future contraceptive applications (27). ScFvs specific for a target have been fused to a toxin to target specific cell death or inhibit protein synthesis (28). Herceptin is a monoclonal antibody that was FDA approved in 1998 and is currently used as a drug to treat specific metastatic breast cancer (29). This is promising for the future of scFvs because monoclonal antibodies are easily made into the scFv format. ScFvs are easier to produce than full length antibodies and are easier to deliver to the site of action. There are scFvs currently in clinical trials (reviewed in, (16).

2.3 Considerations in scFv library design

In order to isolate scFvs for potential drug reagents a logical approach to the design of scFv libraries should be considered. ScFvs contain one binding site made up of two variable domains (Figure 2-2). ScFv variable domains are designed after those found in a natural antibody where one of the variable domains is from the heavy chain and the other is from the light chain (Figure 2-1). Each variable domain contains three complementarity determining regions (CDRs) that make up the hyper-variable loops of the scFv (Figure 2-2). The variable CDRs make up the antigen-binding site of the scFv.

2.3.1 Number of variable CDRs

The ideal scFv library would contain all possible combinations of the twenty amino acids in the CDR regions. Based on the assumption that the average number of interacting residues within the CDRs is approximately twenty (30), a library that contains all 20 amino acids at all twenty positions in the CDR region would consist of 20^{20} possible combinations. In practice it is not possible to screen a scFv library this large. Phage display and the yeast two-hybrid assays can only generate libraries with approximately 10^{10} or 10^{7} members, respectively. This leaves two options. The first option is to create libraries that only sparsely sample the total sequence space and the second option is compromise the diversity by limiting the number of variable sites and/or the number of different amino acids used to vary the sites.

CDRs important in binding a target vary depending on the antibody-antigen complex being analyzed. For most antibody-antigen interaction complexes the heavy chain contributes more than the light chain in binding the antigen (31).

Each CDR contains a different number of amino acids that contribute to antigen binding. For example, CDR1 from the heavy chain (CDR H1) is approximately six amino acids long; however, only four amino acids usually contact the antigen (30,31,32). CDR H2 is usually between eight and ten amino acids long, but only approximately six of these amino acids participate in binding (30,31,32). CDR H3 is found in the middle of the binding pocket of almost all antibody constructs (32). It is the most variable CDR in terms of size, conformation, and amino acid composition (32). CDR H3 is also the most important in determining antibody specificity (32). CDR H3 is usually between seven and nine amino acids in length and all of these amino acids have been shown to be important in antibody-antigen interactions (30,31,32). Light chain CDRs that contribute to antigen binding vary amongst different antibody-antigen complexes. In general, CDR L3 contributes the most to antigen interactions, whereas CDR L2 contributes the least with sometimes only two residues contributing to the antigen interaction (31).

2.3.2 Varied amino acids within the CDRs

An scFv library with all twenty amino acids varying the CDRs is too large to create and screen in yeast. Previous studies have shown that minimal libraries can be produced by varying the CDRs with only a few amino acids (32). These minimal libraries with reduced diversity are able to generate highly specific interactions.

Degenerate codons are used to create random regions (ie CDRs) to generate large and diverse synthetic scFv libraries. These libraries can be used to isolate scFvs that interact specifically with a chosen protein target. Several research groups examined the amino acid composition of antibody CDRs and found that tyrosine occurs in the CDRs of antibodies in an

unusually high abundance (31,33-35). Based on these results another research group generated high affinity fragment antigen binding (Fab) libraries with a minimal number of amino acids within the variable regions (8). Fabs are another type of antibody fragment that are larger than scFvs (36). ScFvs consist of the variable domains from the heavy and light chain, whereas Fabs contain the variable domain and a constant domain from the heavy and light chain (36) (Figure 2-1). Eleven libraries diversified with four amino acids were created and screened against four different proteins (8). Using these minimal amino acid diversity libraries, high affinity Fabs were isolated specific for the targets screened against (8). Some of the Fabs contain binding constants (K_d) with single digit nanomolar affinity for their protein target (8). The library with the largest binding capacity contains CDRs varied with tyrosine, serine, alanine, and aspartate. Tyrosine and serine are present in the CDRs of the three of the eleven libraries that interact with the highest number of proteins (8).

Structural analysis of the binding residues of the Fabs reveal that tyrosine side chains are responsible for mediating most of the antigen recognition (8). The explanation for this phenomenon is that tyrosine has hydrophobic and hydrophilic characteristics allows it to interact with the target in multiple ways and to be accommodated in most environments. These properties enable tyrosine to make highly specific and affinitive contacts with antigen (8). Serine and alanine allow for space and conformational flexibility to facilitate productive contacts between tyrosine and the target protein (8). Large amino acids like tyrosine and tryptophan function well adjacent to smaller amino acids like serine and alanine since they provide space for free rotation, whereas large amino acids next to tyrosine and tryptophan prevent flexibility through steric hindrance (30).

Structural characterization of the Fab-antigen complex reveals that the interaction involves all three heavy chain CDRs; the light chain CDRs contribute minimally to antigen binding (9). The small amino acids have little direct contact with the antigen and instead aid in positioning tyrosine for antigen binding. Based on these results, Fellouse *et al*. constructed a library in which the CDR regions were randomized using only tyrosine and serine (9). One of their Fab libraries varied at ~ 25 positions in only the heavy chain CDRs. With this library, Fabs against a broad range of targets are isolated. This study shows that a Fab library with only tyrosine and serine diversifying the CDRs is sufficient to mediate highly specific antigen recognition (9). Design of these minimal amino acid diversity libraries can be applied to other antibody fragments like scFvs. The goal behind creating such low diversity libraries is to isolate an scFv specific for a target protein. The scFvs can be further improved using affinity maturation to increase binding affinity and obtain a high affinity scFv that can be later used in therapeutics, diagnostics, or research.

2.4 Stability Limitations in ScFv Expression

Designing combinatorial diversity of scFv CDRs is crucial for creating scFv libraries that contain high affinity binders with a reasonable library size. If an scFv is not stably expressed it is not functional; therefore the final diversity of the library is also limited by the expression and solubility of scFvs. The variable heavy and light chain domains are internally stabilized by intra-domain disulphide bonds (Figure 2-2), which are not supported in the reducing environment of the cell (37). These disulfide bonds are highly conserved in all antibodies and stabilize the scFv by connecting the β–sheets of the domains (37). ScFvs are also stabilized by hydrophobic interactions between their domains (5). These inter-domain hydrophobic interactions between the

V_L and V_H domains are not sufficient to keep the two domains stably associated (38). Low stability of the V_H-V_L complex leads to dissociation or 'opening' between the scFv domains, exposing hydrophobic β–sheet residues resulting in inter-molecular associations and aggregation of scFv (Figure 2-3) (39,40). For example, the heavy chain variable domain of one scFv will pair with the light chain variable domain of a different scFv (41).

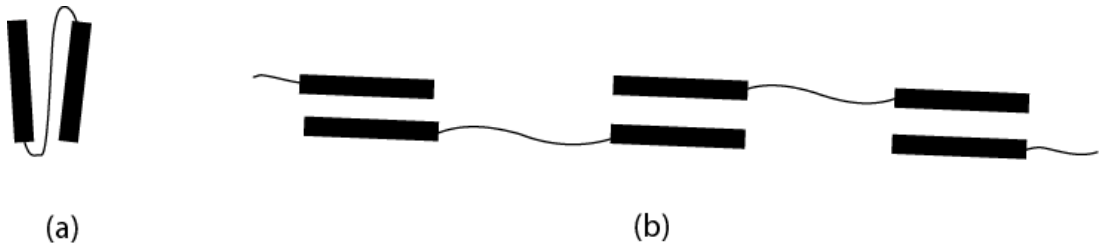


Figure 2-3: ScFv structure and aggregation.
(a) Structure of a properly folded scFv. (b) Aggregated scFvs where the heavy chain of one scFv is interacting with the light chain of a different scFv. Aggregated scFvs are not soluble.

Many different approaches have been attempted to stabilize the scFv structure to prevent aggregation and induce proper folding. Some of these approaches include varying the linker length and composition (42), random mutagenesis of the scaffold and CDRs (5), adding fusion proteins (43), disrupting the hydrophobic patches at the scFv interface (44), mutating the scFv framework (45,46) and co-expressing chaperones (47). Although these attempts to stabilize the scFv structure show improvements in scFv stability they introduce new problems. For example, the formation of dimers or trimers occurs with shorter linker length (48) and steric hindrance prevents proper folding in the presence of some fusion proteins (11).

To be expressed inside a cell, an scFv needs to be able to fold stably in the absence of disulfide bonds and avoid aggregation (49). Selection strategies have been developed in an

attempt to isolate antibody fragments with enhanced stability. Generally this involves mutating a single antibody against a known target and screening for mutations with enhanced stability (37,50). Producing scFvs using this method leads to a predetermined stable scFv framework that limits the user to diversifying only specific CDR positions in fear of disrupting stability. Specific framework amino acids have been shown to stabilize the scFv variable domains in a few different scFvs; however, these specific residues are not sufficient to successfully express and isolate most scFvs in the cytoplasm (5). The amino acids within the variable region of the CDRs contribute to its successful intracellular folding. Certain amino acids can potentially have a negative effect on proper scFv folding.

The proper folding of scFvs can be detected by specific bacterial proteins, which have specificity for antibody framework regions. For example, Protein L from *Peptostreptococcus magnus* binds to most κ-light chains from antibodies (51,52). Protein A from *Staphylococcus aureus* binds to the heavy chain of some antibodies. ScFvs have been successfully purified using both protein A and protein L (53). The binding of an scFv to protein A or protein L is a way to detect and purify properly folded scFvs.

2.5 Inteins and protein cyclization

Cyclizing a protein can increase its stability and solubility (12,14,15). Proteins can be cyclized within cells using a split-intein (54,55). Inteins are self-splicing protein domains that are removed during posttranslational processing (Figure 2-4) (56). Amino acid residues within the extein, adjacent to the splice site and internal residues within the intein control the splicing reactions. Once the intein is spliced out of the protein the flanking proteins (exteins) are joined

by a peptide bond producing the mature protein. Inteins have been found to splice proteins in *cis* as well as in *trans* (57).

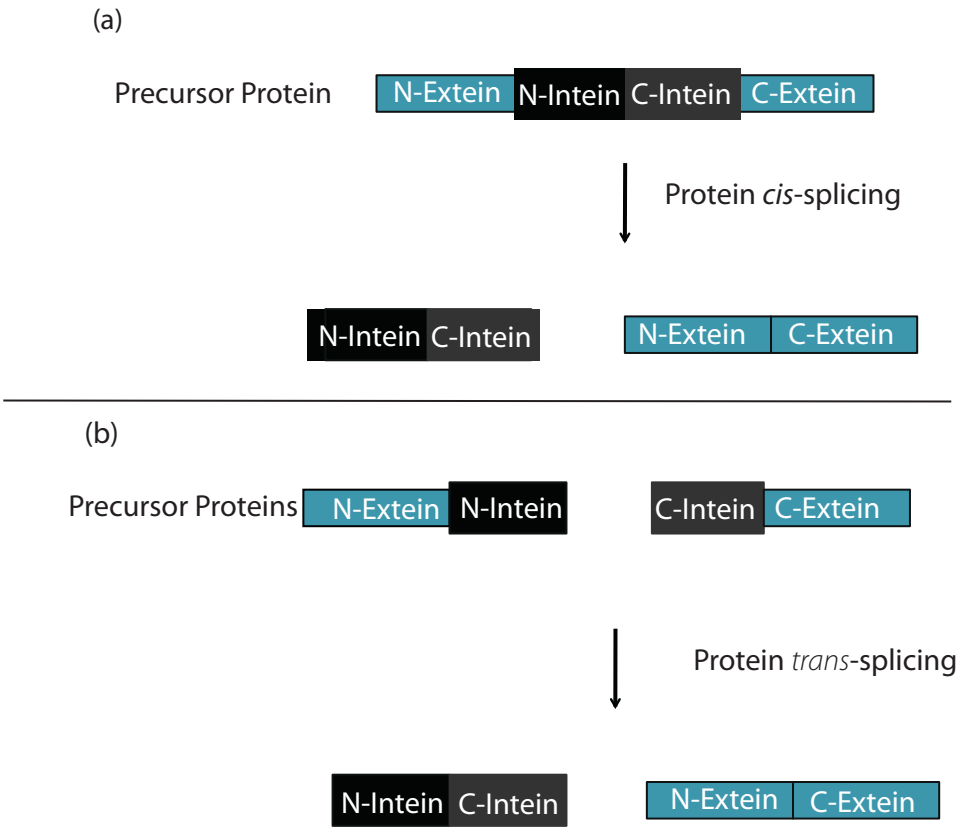


Figure 2-4: *cis* and *trans* splicing of inteins.
(a) *cis* splicing intein; the N terminal and C terminal domains of the intein are joined together and are flanked by the extein sequence to be joined. The intein domains perform the self-excising reaction and are removed from the protein joining the two-extein domains by a peptide bond. (b) *trans*-splicing split-intein; the intein domains are expressed separately with an extein domain. The intein domains catalyze a self-splicing reaction and are excised out of the protein ultimately joining the exteins by a peptide bond.

Trans-splicing allows inteins to be manipulated to cyclize proteins (Figure 2-5). A protein can be cyclized by fusing it between the intein domains with the C-intein domain on the N-terminal end of the protein and the N-Intein domain on the C-terminal end of the protein. The

intein domains catalyze the natural reactions joining the two ends of the extein together ultimately cyclizing the protein.

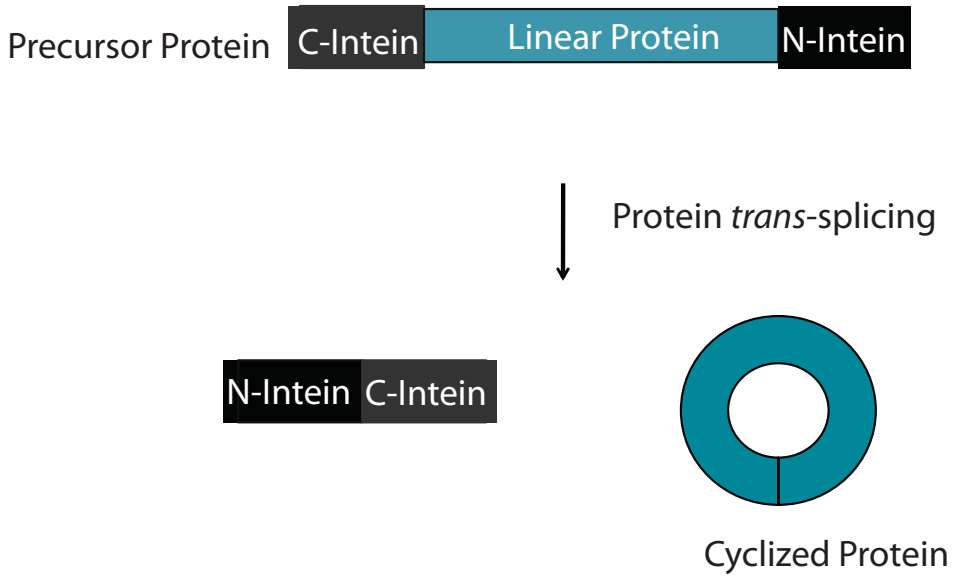


Figure 2-5: Intein mediated protein cyclization.
The protein to be cyclized is place in between the two-intein domains with the N-intein domain and C-intein domain swapped. The intein domains catalyze the self-splicing reaction cyclizing the protein.

The split intein *dna*E from *Synechocystis sp*. (*Ssp*) can be used to cyclize proteins *in vivo* (54). The cyclization of dihydrofolate reductase increases its *in vitro* thermostability and prevents amino-terminal modification and proteolysis (54). Cyclizing the tyrosinase inhibitor pseudostellarin F dramatically increases the amount isolated (54). A different trans-splicing intein dnaE from *Nostoc punctiforme* (*Npu*) can be used to cyclize proteins. Previous studies show that the *Npu* intein is quite tolerant to the amino acid compositions within the target extein (58). Also, the C-intein and N-intein domains from the *Npu* and *Ssp* inteins can be swapped and spliced with >98% efficiency. By using the *Npu* N-intein and *Ssp* C-intein the splicing is more tolerant to different amino acid substitutions within the extein sequence than the *Npu* intein (58).

The hybrid *Ssp-Npu* intein is more useful since it allows more change in all potential proteins to be cyclized.

A recent study shows that the cyclization of two proteins with a split intein stabilizes a binary complex (59). These results provide strong evidence suggesting that cyclization can be used to stabilize scFvs. ScFvs can be constructed in the V_H-linker-V_L (60) or the V_L-linker-V_H (61) orientation, which shows that no structural barriers exist concerning the construction of cyclic scFvs (12,14).

2.6 Screening combinatorial scFv libraries

Isolating functional scFvs is a major problem preventing their use. Many strategies aim to optimize the scFv framework for expression inside the cell; however, it is not clear whether there is a general framework that can support diverse CDRs while retaining intracellular stability. Universal scFv frameworks with randomized CDRs have been used to isolate intracellular scFvs (62). ScFvs can be screened for interactions using *in vitro* display systems such as phage, yeast surface, and ribosome display (4). Phage display is one of the more popular strategies for obtaining intracellular scFvs from a library (6,49,63,64). In phage display the library of scFvs are expressed on the outside of phage by fusion to the phage protein coat. As a result, scFvs isolated by phage display are not optimized for expression inside cells. This technique is limited by the lack of selection pressure for selected antibody fragments to function inside cells. As a result, most of the antibody fragments isolated are not stable and have low expression levels and short half-lives when expressed intracellularly. Recently, Villani *et al* isolated a few scFvs from a library using phage display under reducing conditions, removing any disulphide bonds to increase the chance of the scFvs being stable inside cells (65). ScFvs isolated under reducing

conditions (isolated from bacterial cytoplasm or reduced with DTT) bound to their target with comparable affinity to the scFv in its oxidized state, however, the scFvs could not be isolated with protein L (which is an indicator of correct folding) suggesting that the scFvs were not properly folded or possibly did not need to be properly folded to bind to their target (65). Also, the binding ability was not tested within the environment of a cell, which is a different environment containing obstacles that can prevent the scFv from binding its target (65).

The yeast two-hybrid (Y2H) assay is used to isolate protein interactions within the environment of a eukaryotic cell (66). A mixture of variable scFvs is cloned into Y2H expression vectors, creating a large diverse combinatorial scFv library expressed in yeast (49). Each scFv library member is expressed inside a different yeast cell. A target protein is expressed in the opposite yeast cell mating type from the library (Figure 2-6) (66). ScFv library cells are mated with the cells expressing the target protein, which produces diploid cells. Each diploid cell contains an scFv from the library and the target protein or protein domain. The selection for diploid cells containing an interacting library member and protein domain can be screened for using reporter genes. Reporter genes used in the Y2H are located downstream from a DNA binding domain (66). The target protein domain is tethered to a DNA binding domain. For example, LexA (a DNA binding protein) will bind to its binding site on the DNA, holding the target domain near the promoter of the reporter gene. The scFv or library member is tethered to a transcription activation domain (66). If the scFv library member within the diploid cell interacts with the protein domain the activation domain will activate transcription of the reporter gene. A reporter gene expresses a phenotype that can be observed through cell growth or phenotype.

The Y2H assay is a system that can be used to find an scFv that binds to a protein target of interest, which will select against scFvs that are not expressed or do not fold properly in the

intracellular environment. The Y2H assay still needs to be improved in order to increase the number of library members that can be successfully expressed and screened in the Y2H assay and to decrease the number of false positives.

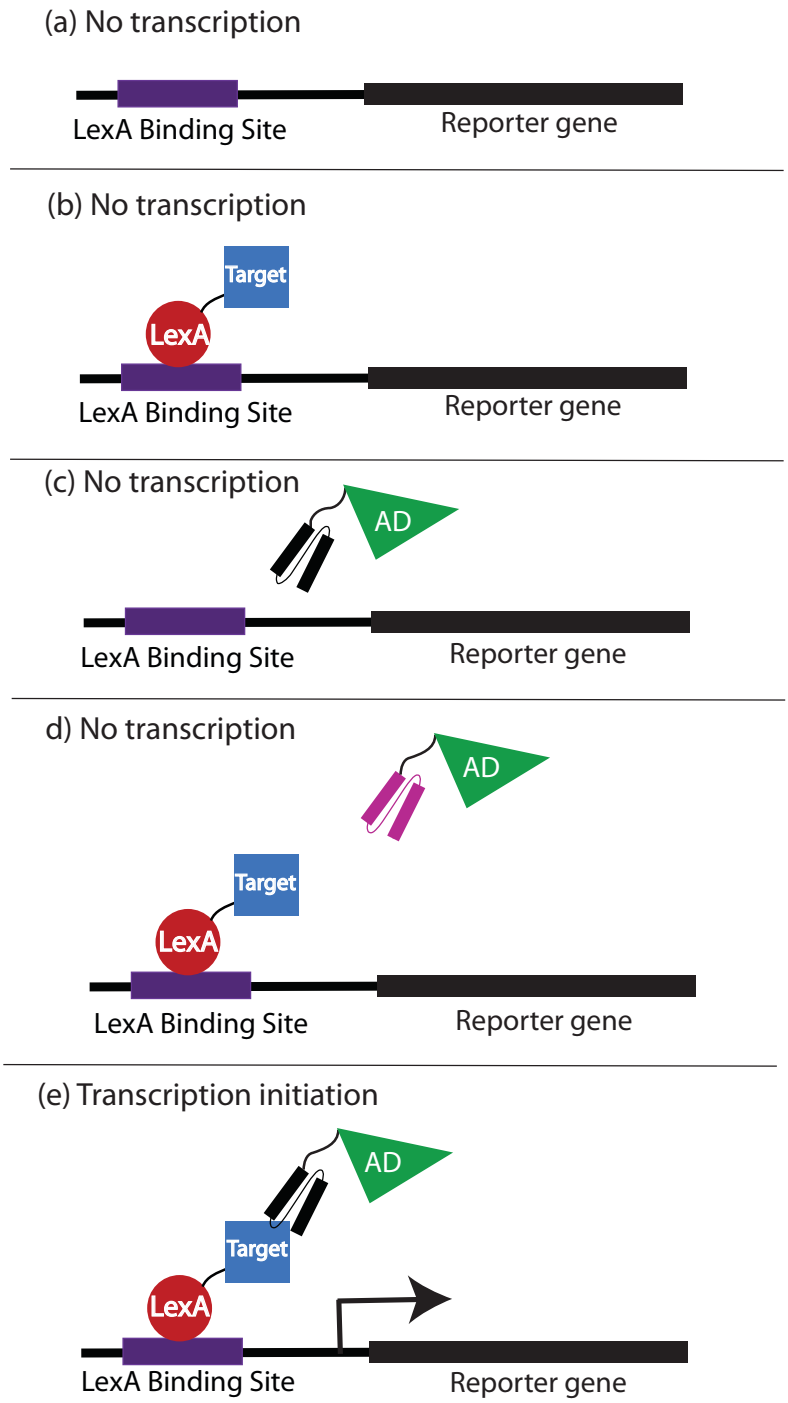


Figure 2-6: ScFv yeast two-hybrid assay.
(a) Without any proteins near the transcription start site there is no transcription initiation. (b) The target protein fused to the LexA DNA binding domain near the transcription start site alone will not initiate transcription. (c) The scFv fused to the activation domain (AD) alone will not activate transcription. (d) If the scFv fusion and target protein do not interact transcription will not be initiated. (e) In the presence of a scFv and target protein that interact transcription of the reporter gene is initiated and the reporter gene is expressed.

Chapter 3 Materials and Methods

3.1 Microorganisms and Media

Media components and chemicals were obtained from VWR (Canada) unless otherwise noted. Oligonucleotides were obtained from Integrated DNA technologies (IDT, USA) (Table 3-1).

3.1.1 *S. cerevisiae* Strains

S. cerevisiae strains used in this study include EY111 and EY93. The general characteristics of the strains are listed below:

EY111 – MATα his3 trp1 ura3::LexA8op-lacZ ade2::URA3-Lex8op-ADE2
leu2::LexA6op-LEU2* derived from EGY48 (67)

EY93 – MATa ura2 his3 trp1 leu2 ade2::URA3 is derived from EGY42 (68)

*Explanation in Glossary section under ::

Standard procedures were used for culturing *S. cerevisiae* as outlined in Geyer *et al*. (69). YPD media consisted of 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (v/v) dextrose; synthetic dropout media (SD): 0.67% (w/v) yeast nitrogen base without amino acids, appropriate amino acid dropout, supplemented with either 2% (v/v) dextrose or 2% (v/v) galactose and 1% (v/v) raffinose. All solid media also contained 2% (w/v) agar. For blue/white screening in *S. cerevisiae* 80 μg/mL of X-gal (Invitrogen, Canada) was added to medium.

Table 3-1: Oligonucleotides used in this study.

#	Name	Length (bp)	Sequence 5' - 3'	Type of Primer	Organism	Description
1	Oligo 1 - Ab	60	CAAGTTCAATTGGTTGAATCTGGTGGTGGTTTGGTT CAACCAGGTGGTTCTTTGAGATTG	scFv Construction	S. *cerevisiae*	Forward primer used to construct the scFv framework
2	Oligo 2 - Ab	54	TCGCGAGAAACCAGAAGCAGCACAAGACAATCTC AAAGAACCACCTGGTTGAAC	scFv Construction	S. *cerevisiae*	Reverse primer used to construct the scFv framework
3	Oligo 3 - Ab	62	GCTGCTTCTGGTTTCTCGCGATACTACGCTGACTTC GTTAAGGGTAGATTCACCATCTCTAG	scFv Construction	S. *cerevisiae*	Forward primer used to construct the scFv framework
4	Oligo 4 - Ab	63	CATTTGCAAGTACAAGGTGTTCTTAGAGTTGTCTCT AGAGATGGTGAATCTACCCTTAACGAA	scFv Construction	S. *cerevisiae*	Reverse primer used to construct the scFv framework
5	Oligo 5 - Ab	56	AACTCTAAGAACACCTTGTACTTGCAAATGAACTCT TTGAGAGCTGAAGACACCGC	scFv Construction	S. *cerevisiae*	Forward primer used to construct the scFv framework
6	Oligo 6 -Ab	57	GATGGTGATCTCGAGTCTAGCACAGTAGTAAACAG CGGTGTCTTCAGCTCTCAAAGA	scFv Construction	S. *cerevisiae*	Reverse primer used to construct the scFv framework
7	Oligo 7 - Ab	58	TGTGCTAGACTCGAGATCACCATCTTCGGTGGTGGT ATGGACGTTTTCGACTACTGGG	scFv Construction	S. *cerevisiae*	Forward primer used to construct the scFv framework
8	Oligo 8 - Ab	52	CAGAAGAAACGGTAACCAAGGTTTGACCCCAGTAG TCGAAAACGTCCATACC	scFv Construction	S. *cerevisiae*	Reverse primer used to construct the scFv framework
9	Oligo 9 - Ab	60	CTTGGTTACCGTTTCTTCTGGTGGTGGTGGTTCTGG TGGTGGTGGTTCTGGTGGTGGTGG	scFv Construction	S. *cerevisiae*	Forward primer used to construct the scFv framework
10	Oligo 10 - Ab	64	GCAGACAAAGAAGATGGAGATTGGGTCATAACGAT GTCAGAACCACCACCACCAGAACCACCAC	scFv Construction	S. *cerevisiae*	Reverse primer used to construct the scFv framework
11	Oligo 11 - Ab	62	ACCCAATCTCCATCTTCTTTGTCTGCTTCTGTTGGTG ACAGAGTTACCATCACCTGTGCTGG	scFv Construction	S. *cerevisiae*	Forward primer used to construct the scFv framework
12	Oligo 12 - Ab	67	GTACCAAGAAACGTAGTTGTAACCACCAACGTCAG AAGAGGTACCAGCACAGGTGATGGTAACTCTG	scFv Construction	S. *cerevisiae*	Reverse primer used to construct the scFv framework
13	Oligo 13 - Ab	63	TTGGTGGTTACAACTACGTTTCTTGGTACCAACAAA AGCCAGGTAAGGCTCCAAAGTTGTTGA	scFv Construction	S. *cerevisiae*	Forward primer used to construct the scFv framework
14	Oligo 14 - Ab	65	CTAGATGGAACACCAGATGGTCTCTTAGAGTCTTCG TAGATCAACAACTTTGGAGCCTTACCTGG	scFv Construction	S. *cerevisiae*	Reverse primer used to construct the scFv framework
15	Oligo 15 - Ab	64	TAAGAGACCATCTGGTGTTCCATCTAGATTCTCTGG TTCTGGTTCTGGTACCGACTTCACCTTG	scFv Construction	S. *cerevisiae*	Forward primer used to construct the scFv framework
16	Oligo 16 - Ab	62	AGGTAGCGAAGTCTTCTGGTTGCAAAGAAGAGATG GTCAAGGTGAAGTCGGTACCAGAACCA	scFv Construction	S. *cerevisiae*	Reverse primer used to construct the scFv framework
17	Oligo 17 - Ab	65	GCAACCAGAAGACTTCGCTACCTACTACTGTTCTTC TACCACCACCAAGTCTACCAGAGTTTTCG	scFv Construction	S. *cerevisiae*	Forward primer used to construct the scFv framework
18	Oligo 18 - Ab	50	TTATCTCTTGATTTCAACCTTGGTACCTTGACCGAA AACTCTGGTAGACTTGGTGGTG	scFv Construction	S. *cerevisiae*	Reverse primer used to construct the scFv framework
19	Ab/pJG4-5.FWD	59	TACCCTTATGATGTGCCAGATTATGCCTCTCCCGAAT TCCAAGTTCAATTGGTTGAATC	Amplification	S. *cerevisiae*	Forward primer used to amplify the scFv to clone into pIL500
20	Ab/pJG4-5.RVS	59	ACTTGACCAAACCTCTGGCGAAGAAGTCCAAAGCT TCTCTTATCTCTTGATTTCAACCT	Amplification	S. *cerevisiae*	Reverse primer used to amplify the scFv to clone into pIL500
21	CDR1-FR2-CDR2.FWD	59	GTGGTGGTTTGGTTCAACCAGGTGGTTCTTTGAGAT TGTCTTGTGCTGCTTCTGGTTTC	Amplification	S. *cerevisiae*	FR2 and CDR H2 to clone into scFv framework
22	CDR1-FR2-CDR2.RVS	59	GTGTTCTTAGAGTTGTCTCTAGAGATGGTGAATCTA CCCTTAACGAAGTCAGCGTAGTA	Amplification	S. *cerevisiae*	FR2 and CDR H2 to clone into scFv framework
23	Oligo-CDR1KMT	86	CTTGTGCTGCTTCTGGTTTCKMTATTKMTKMTKMT KMTATGAACTGGGTTAGACAAGCTCCAGGTAAGGG TTTGGAATGGGTTTCT	CDR Construction	S. *cerevisiae*	Forwared primer used to construct degenerate CDR H1and FR2 from K libaries
24	Oligo-CDR1TMT	86	CTTGTGCTGCTTCTGGTTTCTMTATTTMTTMTTMTT MTATGAACTGGGTTAGACAAGCTCCAGGTAAGGGT TTGGAATGGGTTTCT	CDR Construction	S. *cerevisiae*	Reverse primer used to construct degenerate CDR H1and FR2 from T libaries
25	Oligo-CDR2KMT	73	TTAACGAAGTCAGCGTAGTAAKMAGTAKMACCAK MAKMTGGAKMAATAKMAGAAACCCATTCCAAACC CTTAC	CDR Construction	S. *cerevisiae*	Reverse primer used to construct degenerate CDR H2 and FR2 from K libaries

Table 3-1 (continued)

26	Oligo-CDR2TMT	73	TTAACGAAGTCAGCGTAGTAAKAAGTAKAACCAKAAKATGGAKAAATAKAAGAAACCCATTCCAAACCCTTAC	CDR Construction	S. *cerevisiae*	Reverse primer used to construct degenerate CDR H2 and FR2 from T libaries
27	Oligo-CDR3KMT	61	CTGTTTACTACTGTGCTAGAKMTKMTKMTKMTKMTKMTKMTATCACCATCTTCGGTGGTGG	CDR Construction	S. *cerevisiae*	Forward primer used to construct CDR H3 from K libraries
28	Oligo-CDR3TMT	61	CTGTTTACTACTGTGCTAGATMTTMTTMTTMTTMTTMTTMTATCACCATCTTCGGTGGTGG	CDR Construction	S. *cerevisiae*	Forward primer used to construct CDRH3 from T libraries
29	CDR3.FWD	59	TGTACTTGCAAATGAACTCTTTGAGAGCTGAAGACACCGCTGTTTACTACTGTGCTAGA	CDR Amplification	S. *cerevisiae*	Reverse primer used to amplify CDR H3 to clone into scFv framework
30	CDR3.RVS	59	GTAACCAAGGTTTGACCCCAGTAGTCGAAAACGTCCATACCACCACCGAAGATGGTGAT	CDR Amplification	S. *cerevisiae*	Reverse primer used to amplify CDR H3 to clone into scFv framework
31	TMTL3.RVS	59	CCTGGTACCTTGACCGAAAKATGGAKAAKAAKAAKATTGACAGTAGTAGGTAGCGAAG	CDR Mutagenesis	S. *cerevisiae*	Reverse primer used to diversify CDR L3 in T4 libraries
32	KMTL3.RVS	59	CCTTGGTACCTTGACCGAAAKMTGGAKKAKMAKMAKMTTGACAGTAGTAGGTAGCGAAG	CDR Mutagenesis	S. *cerevisiae*	Reverse primer used to diversify CDR L3 in K4 libraries
33	Ab33.pjG26.RVS	59	ACCTCTGGCGAAGAAGTCCAAAGCTTTCTCTTGATTTCAACCTTGGTACCTTGACCGAA	Extension	S. *cerevisiae*	Reverse primer used to extend scFv sequence after L3 diversification
34	pJG4-5.RVS	59	GGTAGACAAGCCGACAACCTTGATTGGAGACTTGACCAAACCTCTGGCGAAGAAGTCCA	Amplification	S. *cerevisiae*	Reverse primer used to amplify scFv to clone into pIL500
35	P1 Fwd pJG4-5 Chk	20	GGACAGGAGATGCCGATGGA	Amplification	S. *cerevisiae*	Forward primer used to amplify scFv to clone into pIL500
36	P1 IntraVH3-74/pIL500	60	AACTTCTTGTTGGCTAACGGTGCTATTGCTCACGCTTCGCAAGTTCAATTGGTTGAATCT	Amplification	S. *cerevisiae*	Forward primer used to amplify scFV to clone into intein vector
37	P2 IntraL19/Linker	60	ACTTCCTCCGCCGCCAGAACCACCACCACCAGAACC	Extention	S. *cerevisiae*	Reverse primer used to extend scFv to include a second linker
38	P2 Linker/pIL500	60	CAAACCGTATTCAACAGTCAAAATTTCAGTTTCGTAAGACAAACAACTTCCTCCGCCGCC	Amplification	S. *cerevisiae*	Reverse primer used to amplify scFv to clone in to intein vector
39	P1 pET28b/VH3-74	27	CCGAATTCGCAAGTTCAATTGGTTGAA	Amplification	S. *cerevisiae*	Forward primer used to amplify linear scFv to clone into pET28b
40	P2 L19/pET28b	31	TTATCTCGAGTTATCTCTTGATTTCAACCTT	Amplification	S. *cerevisiae*	Reverse primer used to amplify linear scFV to clone into pET28b
41	P1 Intein/EcoR1+1	29	CCGGAATTCGGTTAAGGTTATTGGTAGAA	Amplification	S. *cerevisiae*	Forward prim used to amplify cyclic scFv to clone into pET28b
42	P2 Intein-Npu/XhoI	30	TTATCTCGAGTTAGTTTGGCAAGTTATCAA	Amplification	S. *cerevisiae*	Forward primer used to amplify cyclic scFv to clone into pET28b
43	P1 Hck TyrKin/pEG202	59	GCGGTTGGGGTTATTCGCAACGGCGACTGGCTGGAATTCCTCAAGCTGGAGAAGAAACT	Amplification	S. *cerevisiae*	Tyrosine Kinase domain to clone into pEG202
44	P2 Hck TyrKin/pEG202	59	AATTCGCCCGGAATTAGCTTGGCTGCAGGTCGACTCGAGTCATGGCTGCTGTTGGTACT	Amplification	S. *cerevisiae*	Tyrosine Kinase domain to clone into pEG202

*The name of the oligonucleotide is indicated in the first column. The length of the oligonucleotide in base pairs (bp) is shown in the second column. The sequence is shown in the third column in bases listed from the 5 prime (‘) to the 3 prime (’) end. The organism that the oligonucleotide was used to clone into is indicated in the second last column. The use and type of primer is also shown.

3.1.2 *E. coli*

E. coli strains used in this study include ER2925 (New England Biolabs, Canada), BL21 codon plus (Stratagene, Canada), and XL1Blues (Stratagene). Standard procedures were used for culturing *E. coli* as outlined in (70-72). The LB medium consisted of 1% Tryptone, 0.5% yeast extract, 85.6 mM NaCl, and 1 mM NaOH. For *E. coli* vector selection kanamycin and ampicillin were used at concentrations of 0.05 μg/mL and 0.1 μg/mL respectively when applicable. The solid medium contained 2% (w/v) agar.

3.1.3 Vectors

Vectors used in this study include pEG202 (Figure 3-12) (67), pET28b (Figure 3-13) (Novagen, Canada) and pIL500 (67) (Figure 3-3). pIL500 was supplied by Kris Barreto, a graduate student in Dr. Geyer's lab. The pIL500 plasmid was derived from pJG4-5. A kanamycin gene was cloned in. The Ssp-Npu intein machinery was cloned in between the *Xho*I and *Eco*RI restriction sites flanking an *Nru*I endonuclease restriction site.

3.2 Primer design for yeast amplification and CDR construction

All forward yeast amplification primers were designed by taking 39 base pairs immediately before the clone site from the vector sequence and 20 base pairs from the beginning of the DNA insert. All reverse yeast amplification primers were designed by taking 39 base pairs immediately after the clone site from the vector sequence and 20 base pairs from the end of the DNA insert.

The CDR construction primers are more variable in composition and construction because they contain degenerate codons. For CDR H1 primers, 20 base pairs were taken from the scFv framework immediately before the *Nru*I restriction site followed by 18 base pairs

comprising 6 degenerate codons followed by 48 base pairs to create the majority of framework 2 (FR2) sequence between CDR H1 and CDR H2 (Figure 3-7).

The CDR H2 primer contained 20 base pairs from the end of FR2, 30 base pairs comprising five degenerate codons followed by 23 base pairs from the scFv immediately after the *Nru*I restriction site.

The CDR H3 site was made from 20 base pairs from the scFv framework immediately before the *Xho*I site, 21 base pairs comprising seven degenerate codons followed by 20 base pairs from the scFv framework immediately after the *Xho*I site.

The forward *E. coli* amplification primers were created by taking the first nine base pairs from pET28b immediately after the *Xho*I site and the last 17 or 18 base pairs from the beginning of the scFv or I_C domain depending if protein was linear or cyclic respectively. The first nine base pairs include two extra base pairs in front of the *Eco*RI site and one extra base after to ensure the insert will be cloned in-frame.

The reverse E. *coli* amplification primers were created from 13 base pairs from pET28b before the *Eco*RI site and 17 base pairs from the end of the scFv or I_N domain depending on if the protein was linear or cyclic. The 13 base pair region includes three or four extra bases before the *Xho*I site, the *Xho*I site and 1-3 base pairs after the site to ensure the scFv is in the correct reading frame.

3.3 Construction of ScFv and ScFv Library Design

3.3.1 ScFv Framework Gene Construction

A synthetic gene encoding the scFv framework (Figure 3-1) was constructed using codons optimized for expression in *S. cerevisiae*. Amino acid sequences of the heavy and light

chain were designed based on scFvs previously reported by Tanka *et al* (8,73,74). The region spanning the first and second CDR of the heavy chain was replaced with an *NruI* endonuclease restriction site to allow cloning in of random amino acids comprising CDR H1 and CDR H2 of the heavy chain (Figure 3-2). CDR H3, the third CDR on the heavy chain was replaced by a *XhoI* restriction endonuclease site. The light chain CDRs were fixed based on an anti-β-galactosidase scFv reported by Martineau *et al* (45).

The program GeneDesign (75) was used to design the eighteen overlapping oligonucleotides (Table 3-1) that were used to construct the synthetic scFv framework gene. The eighteen oligonucleotides (Oligo1-Ab to Oligo18-Ab) were mixed together (0.2 ng/μL of each) with 1X Platinum HIFI taq polymerase buffer (60 mM Tris-SO_4 (pH 8.9), 18 mM $(NH_4)_2SO_4$ (Invitrogen), 0.2 mM dNTPs, 2 mM $MgSO_4$, and 1.0 unit Platinum HIFI taq polymerase (Invitrogen). The reaction mixture was placed in a thermocycler under the following conditions: 94 °C for 2 minutes, [94 °C for 30 seconds, 56 °C for 30 seconds, 68 °C for 1 minute](30 cycles), and 68 °C for 10 minutes. The full-length gene product was amplified in a second PCR reaction (Figure 3-1) where two microlitres of PCR #1 was mixed with 0.2 μM Ab-pJG4-5.Fwd primer (#19 Table 3-1) and 0.2 μM of Ab-pJG4-5.Rvs primer (#20 Table 3-1) using the conditions described above.

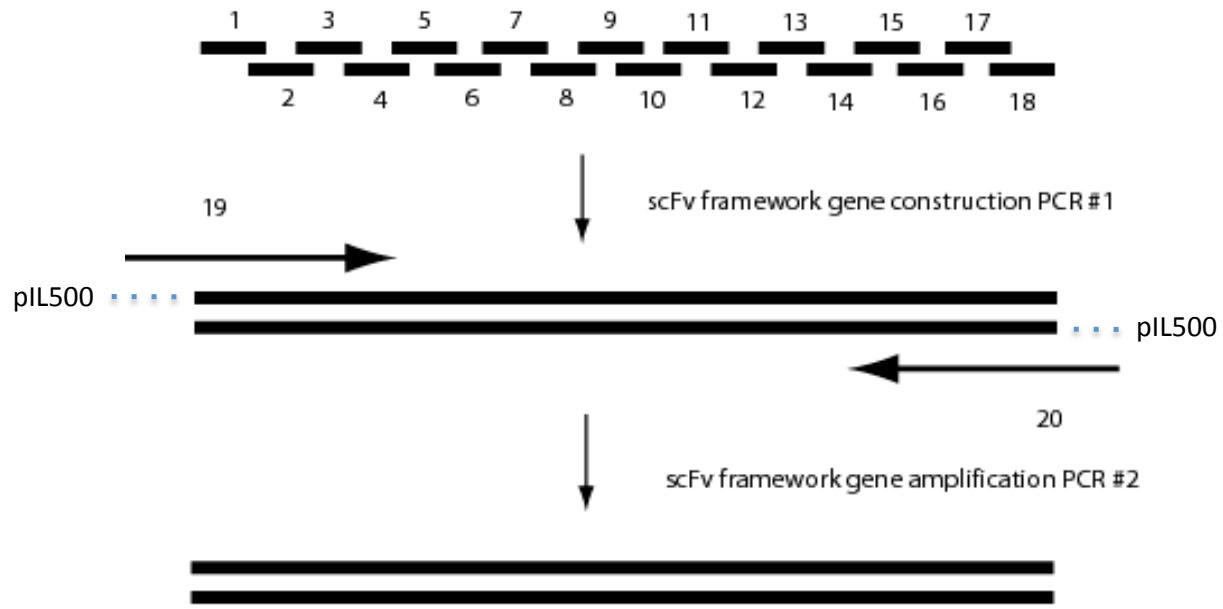


Figure 3-1: ScFv Framework Gene Construction.
The scFv framework was created using eighteen overlapping oligonucleotides, which were amplified in PCR reaction #1. The scFv scaffold was PCR amplified in PCR reaction #2 using primers with complementarity to pIL500 (….). The amplified scFv scaffold was cloned into pIL500. The numbers indicate the primer used at each step and are shown in Table 3-1.

The scFv framework heavy chain sequence contains a *Nru*I site (in place of CDRH1, framework 2 (FR2) and CDR H2) flanked by framework 1 (FR1) and framework 3 (FR3) (Figure 3-2). A *Xho*I site (in place of CDR H3) flanked by FR3 and framework 4 (FR4). A linker peptide fuses the heavy chain to the light chain.

Figure 3-2: ScFv Framework.
The scFv framework is an incomplete scFv. The CDR H1, FR2, and CDR H2 have been replaced by a *Nru*I restriction site. The CDR H3 has been replaced with a *Xho*I site. The light chain CDRs are in place but they are not variable. The black segments represent the framework sequences (FR). The thin line indicates the linker peptide and the grey segments show the light chain CDRs.

3.3.2 Cloning the ScFv Framework into Yeast Expression Vector

pIL500 (Figure 3-3) was isolated from *E. coli* using QIAprep Spin Miniprep kit (Qiagen, Canada) according to the manufacturer's instructions. pIL500 was digested with 0.5 units of *Eco*RI endonuclease (New England Biolabs), 0.5 units of *XhoI* endonuclease (Fermentas, Canada) and 1X y+/Tango buffer (Fermentas) in a 20 µL reaction volume. The reaction was incubated at 37 °C overnight. The PCR amplified scFv framework was cloned into the *Eco*RI and *Xho*I digested pIL500 (Figure 3-4) using homologous recombination by co-transformation into yeast strain EY93 as described by Schiestl *et al* (76).

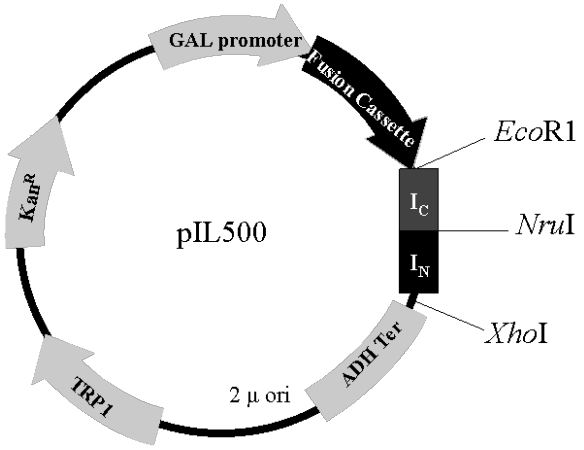


Figure 3-3: Yeast Protein Expression Vector pIL500.
pIL500 contains a galactose inducible promoter (GAL promoter), an alcohol dehydrogenase terminator (ADH Ter) and a 2 micron origin of replication (2 µ ori). The fusion cassette consists of a nuclear localization signal, which localizes the expressed protein to the nucleus, a B42 activation domain, which can activate transcription in the yeast two-hybrid system, and a hemagglutin or HA tag which can be used to purify the expressed protein. pIL500 contains an I_C and I_N domain from the Ssp-Npu intein which catalyzes the cyclization. Linear proteins cloned into the *Eco*RI and *Xho*I restriction sites will remove the intein machinery and will be expressed linearly. pIL500 contains the first gene involved in tryptophan synthesis (TRP1) allowing cells containing this plasmid to grow on media lacking tryptophan. It also contains a gene for kanamycin resistance allowing cells containing this plasmid to grow in media containing kanamycin.

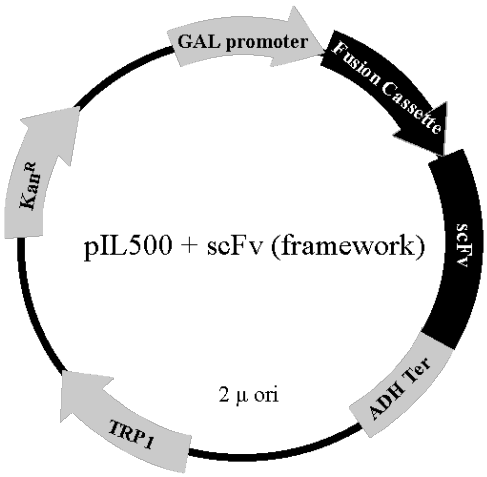


Figure 3-4: Linear scFv (framework) cloned into pIL500.
This figure depicts the scFv framework and full size scFv. They will look almost the same cloned into pIL500 with the exception that the full size scFv will be a little larger because it will contain all three heavy chain CDRs. The scFv (framework) is cloned into the restriction site *EcoRI* and *Xho*I. When pIL500 is digested with *Eco*RI and *Xho*I, the *Nru*I restriction site as well as the Ssp-Npu split intein machinery for protein cyclization will be removed (Figure 3-3). This plasmid contains the same promoter, terminator, fusion cassette and markers as pIL500.

3.3.3 CDR Library Oligonucleotides

The scFv CDRs were created by amplifying degenerate oligonucleotides. The degenerate oligonucleotides were cloned into the scFv framework using homologous recombination (76). Combinatorial libraries consisting of two amino acids, Tyrosine (TAT codon) and Serine (TCT codon) are referred to as the "T" libraries. The two amino acid libraries or T libraries were constructed using TMT degenerate codons, where T = thymine and M = adenine or cytosine. The T libraries were designated as T3 or T4 depending on the number of CDRs varied within the structure (Table 3-2). The T3 libraries have heavy chains CDR H1, CDR H2, and CDR H3 varied with either tyrosine or serine (Figure 3-5).

CDR-H1

28	29	30	31	32	33
X	I	X	X	X	X

CDR-H2

50	51	52	52a	53	54	55	56	57	58
X	I	X	P	X	X	G	X	T	X

CDR-H3

95	96	97	98	99	100	100a
X	X	X	X	X	X	X

CDR-L3

90	91	92	93	94	95	96
Q	X	X	X	X	P	X

Figure 3-5: Positions of diversified complementarity determining regions (CDRs).
The names of the CDRs are listed above the tables and the positions are labeled with numbers corresponding to Kabat (77). The heavy chain positions will be varied in all the libraries. The light chain positions will be varied in some libraries. The letters under the numbers refer to the amino acid in that position in single letter amino acid code. X denotes the variable positions.

The T4 libraries have heavy chains CDR H1, CDR H2, CDR H3 and light chain CDR L3 varied with either serine or tyrosine. Cyclic and linear versions of T3 and T4 were created; to distinguish between them they are referred to as cT3 and cT4 for cyclic.

Table 3-2: Explanation of the library names**.**

Library Name	Amino Acids	CDRs Varied	Configuration
LT3	Tyrosine, Serine	H1, H2, H3	Linear
cT3	Tyrosine, Serine	H1, H2, H3	Cyclic
LT4	Tyrosine, Serine	H1, H2, H3, L3	Linear
cT4	Tyrosine, Serine	H1, H2, H3, L3	Cyclic
LK3	Tyrosine, Serine, Alanine, Aspartate	H1, H2, H3	Linear
cK3	Tyrosine, Serine, Alanine, Aspartate	H1, H2, H3	Cyclic
LK4	Tyrosine, Serine, Alanine, Aspartate	H1, H2, H3, L3	Linear
cK4	Tyrosine, Serine, Alanine, Aspartate	H1, H2, H3, L3	Cyclic

*The name of each library contains information that represents the configuration, the number of CDRs varied, and the number of amino acids used to vary the CDRs. The configuration is represented in the library name by ‘C’ for cyclic or ‘L’ for linear The number of amino acids used to vary the CDRs is represented in the library name by T or K; ‘T’ represents two amino acids and ‘K’ represents four amino acids. The number of CDRs that were varied is represented in the library name as ‘3’ or ‘4’. The CDRs that were varied are indicated in the third column, where H represents heavy chain, L3 represents light; chain CDR 3.

Combinatorial libraries consisting of amino acids tyrosine (TAT codon), serine (TCT codon), aspartate (GAT codon), and alanine (GCT) were referred to as the "K" libraries. The K libraries were designated as K3 or K4 depending on the number of CDRs varied within the scFv structure (Table 3-2). The K3 libraries have heavy chains CDR H1, CDR H2, and CDR H3 varied with tyrosine, serine, aspartate, or alanine (Figure 3-5). The K4 libraries have heavy chains CDR H1, CDR H2, CDR H3 and light chain CDR L3 varied with serine, tyrosine, aspartate or alanine. Cyclized and linear scFv library versions of both K3 and K4 were created; to distinguish between them, they are referred to as cK3 or cK4 for cyclic, and LK3 or LK4 for linear (Table 3-2).

CDRs containing degenerate oligonucleotides were constructed in the following reaction: 0.2 mM dNTPs, 1X PCR Buffer, 2 units of HiFi Taq polymerase. The template DNA was constructed and amplified using the primers indicated in Table 3-3. The reaction mixture was incubated in a thermocycler at 95 °C for 1 minute, [95 °C for 30 seconds, 52 °C for 30 seconds, 68 °C for 30 seconds] (30 cycles).

Table 3-3: Primers used in the construction of the CDRs.

CDRs	Primers Used*	Primer
		Concentration (μM)
T, H1 and H2	Oligo-CDR1TMT	0.02
	Oligo-CDR2TMT	0.02
	CDR1-FR2-CDR2.FWD	0.2
	CDR1-FR2-CDR2.RVS	0.2
K, H1 and H2	Oligo-CDR1KMT	0.02
	Oligo-CDR2KMT	0.02
	CDR1-FR2-CDR2.FWD	0.2
	CDRR1-F2-CDR2.RVS	0.2
K, H3	KMT CDR3	0.2
	CDR3.FWD	1
	CDR3.RVS	1
T, H3	TMT CDR3	0.2
	CDR3.FWD	1
	CDR3.RVS	1

*DNA sequences of primers are shown in Table 3-1. The CDRs being referred to are listed in the first column, K indicates that it is from the four amino acid library and T indicates it is from the two amino acid library; H1, H2 and H3 represent CDR 1, 2, or 3 from the heavy chain. The oligonucleotides or primers used are listed in the second column Primer concentration indicates the micromolar (μM) concentration of the oligonucleotides used in the total reaction mix.

3.3.4 Cloning Degenerate Heavy Chain CDR H3 into ScFv Framework

The degenerate CDR H3 was PCR amplified as described above (Table 3-1, Table 3-3) The scFv framework in pIL500 was digested with *Xho*I in a 100 μL reaction consisting of 1X NEBuffer 2 (New England Biolabs), 90 μL of scFv framework in pIL500, and 1 unit of *Xho*I. The reaction was incubated for 24 hours at 37 °C. The digested scFv framework in pIL500 was gel purified in a 0.5% agarose gel using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. The degenerate CDR H3 regions for the K and T libraries were cloned into the scFv framework in pIL500 using the lithium acetate transformation protocol (Figure 3-6) (76).

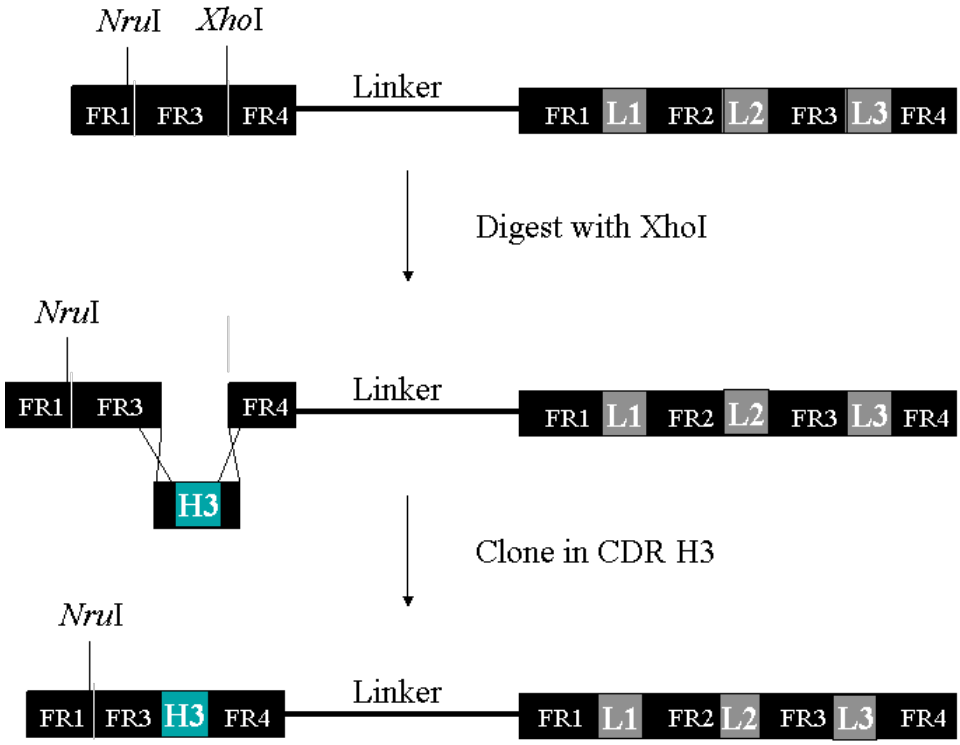


Figure 3-6: Cloning CDR H3 into the scFv framework.
The scFv framework refers to an incomplete scFv framework with the CDRs of the heavy chain and framework 2 (FR2) replaced with two restriction sites. CDR H3 was replaced with a *Xho*I restriction site. To clone CDR H3 into the scFv framework in pIL500, the plasmid was digested with *Xho*I. The CDR H3 was PCR amplified with degenerate codons to create seven variable regions. The amplified CDR H3 was cloned into the scFv framework in pI500 using yeast homologous recombination. FR stands for framework, H stands for heavy chain, and L represents light chain.

In addition to the CDR H3 created in the PCR reaction, the scFv framework contains thirteen amino acids from the ß-galactosidase scFv CDR H3 (73) (Figure 3-7). Human CDR H3 sequences are typically 15 to 20 amino acids long (33). This part of the CDR H3 was included to mimic the average size of the human CDR H3 without increasing the diversity of the library.

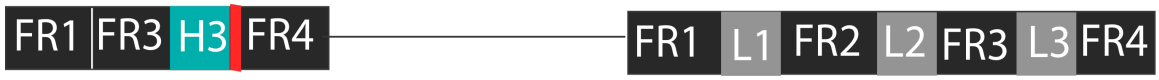


Figure 3-7: Location of ß-galactosidase scFv CDR H3.
The scFv framework sequence includes thirteen amino acids from the ß-galactosidase scFv CDR H3 (73). The location of the additional sequence is shown in red. FR stands for framework, H stands for heavy chain, and L represents light chain.

3.3.5 Cloning Degenerate Heavy chain CDR H1 and CDR H2

To clone degenerate CDR H1 and CDR H2 into the scFv framework containing CDR H3 in pIL500 the CDR H1 and CDR H2 regions were PCR amplified using the PCR reaction described above (Table 3-2, Table 3-3). The scFv framework (+ CDR H3) in pIL500 was digested with *NruI* (New England Biolabs) in a 100 µL reaction volume consisting of 1X NEBuffer 3 (New England Biolabs), 90 µL of the scFv framework (+ CDR H3) in pIL500, and 1 unit of *Nru*I restriction endonuclease. The reaction was incubated at 37 °C for 24 hours. The digested plasmid was gel purified in a 0.5% agarose gel with QIAquick Gel Extraction Kit according to manufacturer's instructions. CDRs H1 and H2 were cloned in using the lithium acetate transformation (Figure 3-8) (76), giving rise to K3 and T3 libraries.

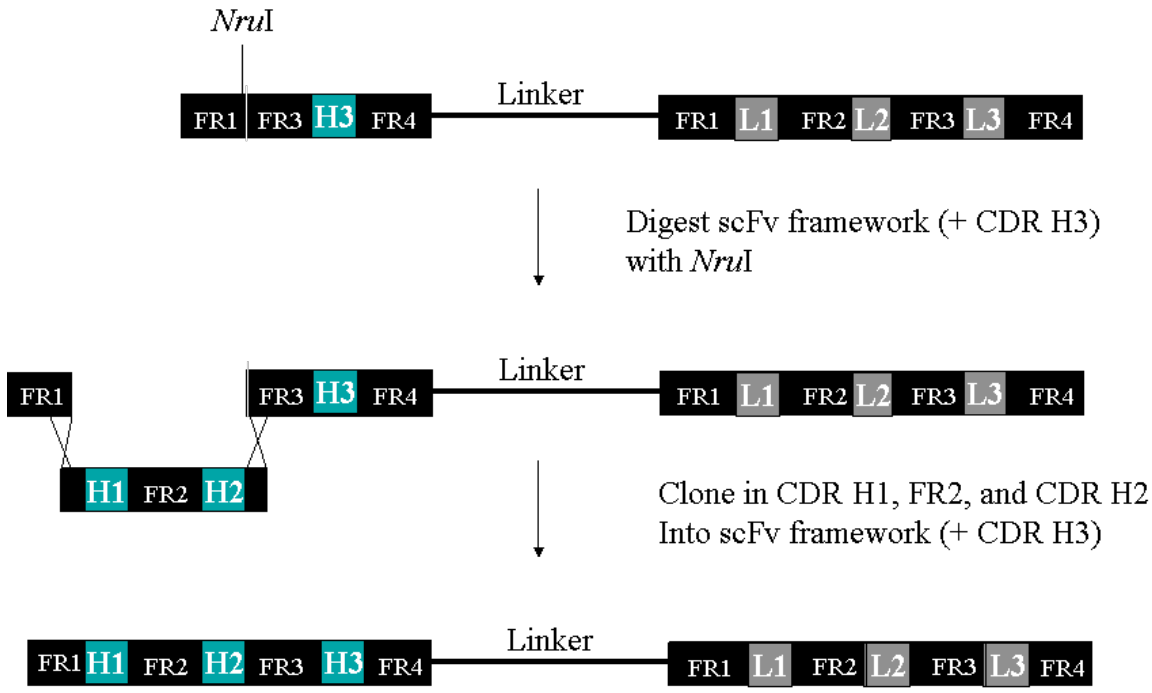


Figure 3-8: Cloning in CDR H1, FR2, and CDR H2.
The scFv framework (+ CDR H3) refers to an incomplete scFv framework, which contains a *Nru*I site in place of the CDR H1, FR2, CDR H2. To clone the insert DNA, CDR H1, FR2 and CDR H2 into the scFv framework (+ CDR H3) the framework in pIL500 was digested with *Nru*I. The insert DNA was PCR amplified and cloned in using homologous recombination in yeast. FR stands for framework, H stands for heavy chain, and L represents light chain.

3.3.5 Mutating Light chain CDR L3 in ScFv

To introduce variation at the light chain CDR L3 into the K3 and T3 libraries, a degenerate oligonucleotide was used as a PCR primer (Table 3-3) to amplify the entire scFv framework in pIL500 containing degenerate CDRs H1-3 (Figure 3-9). The following PCR reaction conditions were used: 1X PCR Buffer (50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$), 0.2 mM dNTPs, 1 µL scFv framework in pIL500 (K3 or T3), 0.6 µM P1 pJG4-5 chK (#35 Table 3-1), 0.2 µM TMT (or KMT)L3.RVS (#31 and #32 Table 3-1), 0.2 µM Ab33.pJG26.RVS (#33 Table 3-1), 0.2 µM pJG4-5.RVS (#34 Table 3-1), 1 µL Taq polymerase (78). The reaction mixture was subjected to a thermocycler under the following conditions: 95

°C for 2 minutes, [95 °C for 30 seconds, 55 °C for 30 second, 72 °C for 30 seconds] (30 cycles), and 72 °C for 10 minutes. The PCR product was cloned into pIL500 digested with *Eco*RI and *Xho*I using lithium acetate transformation (76) giving rise to plasmids expressing K4 and T4 libraries

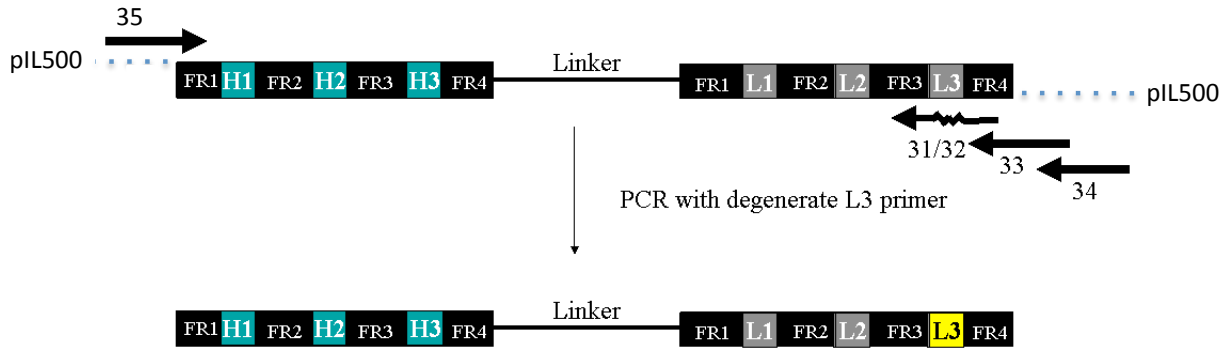


Figure 3-9: Mutational PCR of light chain CDR L3.
The libraries that have four diversified CDRs were derived from the libraries with three diversified CDRs. The libraries with three variable heavy chain CDRs were PCR amplified with a mutational primer containing degenerate codons to diversify the CDR L3. Three different reverse primers were used to obtain enough complementarity to clone the PCR back into pIL500 (….). The primers are represented by the bold arrows and the numbers located under the bold arrows indicates the primers used. The diversified CDR L3 is indicated in yellow. The diversified heavy chain CDRs are indicated in green.

3.3.6 Cyclization of the ScFv Library

pIL500 was digested with 1 unit of *Nru*I and 1X NEBuffer 2 in a 100 µL reaction. The reaction was incubated at 37 ˚C for 24 hours. Plasmids expressing complete libraries with diversified heavy chain CDRs H1-3 (LT3 and LK3) and diversified light chain CDR L3 (LT4 and LK4) were PCR amplified in two PCR reactions with primers containing overlapping complementary sequences to the *Ssp-Npu* split intein (Figure 3-10). In the first PCR reaction the primers were used to add the I_C domain to the beginning of the scFv library and a second linker

to aid in cyclization. The following 100 µL PCR reaction was used: 1 X PCR Buffer, 0.2 mM dNTPs, 1 µL T3, T4, K3, or K4 scFv library plasmid, 0.2 µM P1 IntraVH3-74/pIL500 (#36 Table 3-1), 0.2 µM P2 IntraL19/Linker (#37 Table 3-1), and 1 µL Taq polymerase. The PCR reaction was amplified in a thermocycler under the following conditions: 95 °C for 2 minutes, [95 °C for 30 seconds, 50 °C for 30 seconds, 72 °C for 2 minute] (30 cycles), and 72 °C for 7 minutes. The second PCR reaction was used to add the I_N domain to the end of the second linker.

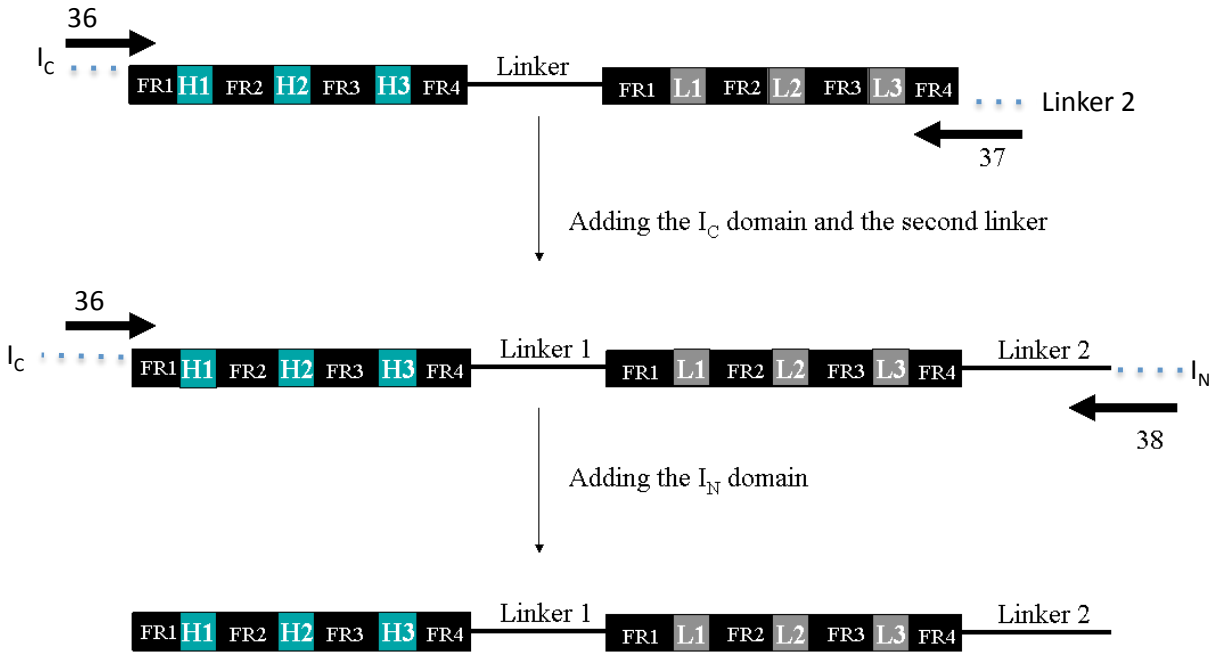


Figure 3-10: PCR amplification to cyclize the scFv libraries.
All the scFv libraries were PCR amplified in two different PCR reactions. In the first reaction, a region complementary to the I_C domain, as well as a second linker were added to the end of the light chain. In the second a region complementary to the I_N domain was added. The I_C and I_N domains will catalyze a cyclization reaction, which will cyclize the scFv with the second linker.

The scFv libraries were cloned into the *Nr*uI restriction site of pIL500 between the *Ssp-Npu* split intein to allow for protein cyclization. The following 100 µL PCR reaction was used:

1X PCR Buffer, 0.2 mM dNTPs, 1 µL LT3, LT4, LK3, or LK4 scFv library plasmid, 0.2 µM P1 IntraVH3-74/pIL500 (#36 Table 3-1), 0.2 µM P2 Linker/pIL500 (#38 Table 3-1), and 1 µL Taq polymerase. The PCR reaction was amplified in a thermocycler under the following conditions: 95 °C for 2 minutes, [95 °C for 30 seconds, 50 °C for 30 seconds, 72 °C for 2 minute] (30 cycles), and 72 °C for 7 minutes. The PCR product was cloned into pIL500 digested with *Nru*I using the lithium acetate transformation in yeast (Figure 3-11) (76). This gives rise to the cT3, cT4, cK3, and cK4 libraries. Fifty members from each library were sent to NRC-PBI to be sequenced to determine the percent of functional scFvs and to confirm library diversity.

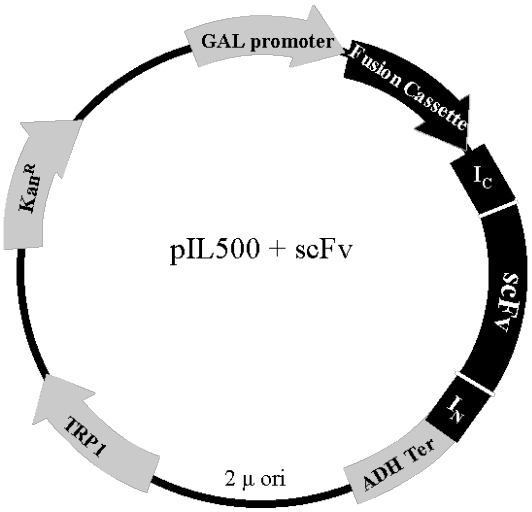


Figure 3-11: Cyclic scFv cloned into yeast expression vector.
The scFv libraries were cloned into the *Nru*I restriction site of pIL500 in between the I_C and I_N domain of the Ssp-Npu split intein. The I_C and I_N domains of the Ssp-Npu split intein will catalyze a cyclization reaction to cyclize the scFv and second linker. This plasmid contains the same promoter, terminator, fusion cassette and markers as pIL500.

3.4 ScFv Yeast Two-Hybrid Interaction Mating Assay

3.4.1 Protein Domains used in Diploid Construction

To ensure the binding capacity of the libraries was accurately sampled, a number of different protein domains were used. To avoid bias in favor of one library over another the protein domains used were diverse in composition and size. The breakpoint cluster region – ableson kinase (BCR-ABL) protein domains (cc1-72, SH2, SH3, PxxP, FABD and Y177) used in this study were supplied by Alberto Aparicio, a graduate student in Dr. Geyer's lab. Cc1-72 is the coiled coil domain from the BCR portion of BCR-ABL, amino acid residues 1- 72. The SH2 and SH3 domains are from the ABL portion of BCR-ABL. The PxxP is a poly-proline region of BCR-ABL that is rich in proline residues. The FABD is a domain of BCR-ABL that binds f-actin. Y177 is a phospho-tyrosine peptide from BCR-ABL.

Hemopoietic cell kinase (Hck) tyrosine kinase domain was also used as a protein domain in this study. Hck was cloned out of pESC-Hck (A gift from Smithgall) and amplified by PCR in the following 100 µL reaction: 1 X PCR Buffer, 0.2 mM dNTPs, 0.2 µM P1 Hck TyrKin/pEG202 (#43 Table 3-1), 0.2 µM P2 Hck TyrKin/pEG202 (#44 Table 3-1), 1 µL pESC-Hck plasmid, and 1 µL Taq polymerase. The PCR reaction was amplified in a thermocycler under the following conditions: 95 °C for 5 minutes, [95 °C for 30 seconds, 55 °C for 30 seconds, 72 °C for 2 minute] (25 cycles), and 72 °C for 7 minutes. The amplified DNA from each protein domain was cloned into the yeast vector pEG202 (Figure 3-12) digested with *Eco*RI and *Xho*I the using lithium acetate transformation (76).

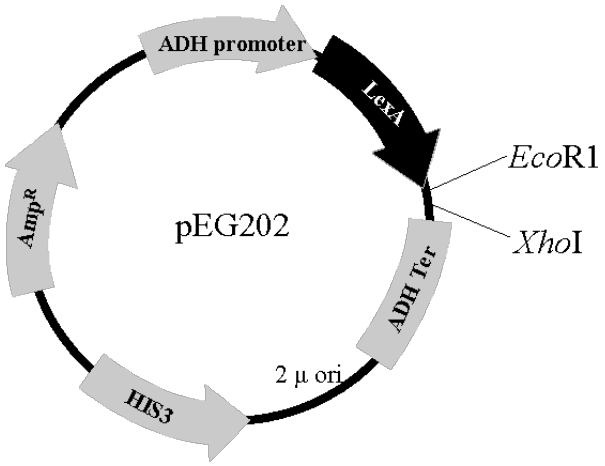


Figure 3-12: pEG202 yeast expression vector.
pEG202 contains a constitutively active alcohol dehydrogenase (ADH) promoter, an alcohol dehyrdrogenase terminator (ADH Ter) and a 2 micron origin of replication (2 μ ori). The protein domains cloned into pEG202 were cloned into the *Eco*RI site and the *Xho*I site. When the protein domains are expressed they are tethered to a LexA binding domain (LexA) which will bind specifically to LexA binding sites on DNA within the yeast cell. pEG202 contains the third gene involved in histidine synthesis (HIS3) allowing cells containing this plasmid to grow on media lacking histidine. It also contains a gene for ampicillin resistance allowing cells containing this plasmid to grow in media containing ampicillin.

3.4.2 Diploid construction

Diploids were constructed using five out of the seven different protein domains (or baits) described above and one of the scFv libraries (Table 3-4).

Table 3-4: Baits Used in each Screen.

Number of Library members Screened	2.6×10^5	1×10^6	6×10^6	12×10^6	20×10^6
Protein Domains Used	BCR-ABL SH2	BCR-ABL SH2	BCR-ABL SH2	BCR-ABL SH2	BCR-ABL SH2
	BCR-ABL SH3	BCR-ABL SH3	BCR-ABL SH3	BCR-ABL cc1-72	BCR-ABL cc1-72
	BCR-ABL cc1-72	BCR-ABL cc1-72	BCR-ABL cc1-72	BCR-ABL Y177	BCR-ABL Y177
	BCR-ABL Y177	BCR-ABL Y177	BCR-ABL Y177	BCR-ABL PxxP	BCR-ABL PxxP
	HCK SH1	HCK SH1	HCK SH1	BCR-ABL FABD	BCR-ABL FABD
Libraries Screened	K3	K3	K3	K3	K3
	K4	K4	K4	K4	cK3
	cK4	cK4	cK4	cK4	K4
	T3	T3	T3	T3	cK4
	T4	T4	T4	cT3	T3
	cT4	cT4	cT4	T4	cT3
				cT4	T4
					cT4

*The number of scFv library members used in each screen is indicated in the top row. The protein domains from the Break point cluster region-Abelson kinase protein (BCR-ABL) and hemotopoietic cell kinase (HCK) are listed below. Protein domains used from BCR-ABL include src homology domain 2 (SH2), src homology domain 3 (SH3), residues 1-72 of the coiled coil oligomerization domain (cc1-72), tyrosine residue 177 + 10 flanking residues (Y177), F-actin binding domain (FABD), and the poly proline domain (PxxP). The src homology kinase domain was also used from HCK.

The protein domains expressed in yeast strain EY111 were grown to an optical density of 0.6-0.9 at 600 nm in SD H^- media. The scFv library cells expressed in yeast strain EY93 were grown to an optical density of 0.6-0.9 at 600 nm in SD W^- liquid media. Both cells were centrifuged at 3220 x g for 5 minutes at room temperature, washed in 1 mL of YPDA+ media and mixed at a 1:20 prey to bait ratio. The cells were gently mixed and incubated at room temperature for five minutes and spread onto YPDA+ media plates. The cells were grown 24 hours at 30 °C, scraped off the plate, washed with 40 mL of ddH_20 and resuspended in the pellet volume of glycerol freeze down solution and stored at – 80 °C. To count the diploids the cells were plated on SD H^-W^- plates and counted after two days of incubation at 30 °C.

3.4.3 Yeast Two-Hybrid Interaction Assay

Diploids were plated on to SG/R $H^-W^-L^-$ media plates and incubated at 30 °C for seven days. The plates were replica plated to SG/R $H^-W^-A^-$ -X-gal media plates with velvet and grown for up for five days at 30 °C. The number of colonies were counted and recorded.

3.4.4 Quantitative Analysis for the Screening Results

Each library was screened in quintuplicate to measure error and reproducibility. The set of data from each screen was averaged to get the number of positive hits from each screen. The number of positive hits from each screen was normalized to compare the cyclic and linear libraries by calculating the ratio of positive hits obtained from the cyclic library divided by the number of positive hits obtained from the linear library. The number of positive hits from each screen was normalized to compare the number of diversified CDRs by dividing the number of positive hits obtained by the three CDR libraries by positive hits obtained from the four CDR libraries. The number of positive hits from each screen was also normalized to compare the number of amino acids used to diversify the CDRs by dividing the positive hits obtained from the four amino acid libraries by the two amino acid libraries. The ratio of positive hits from each screen was plotted against the number of scFv library members screened to determine the trend of each parameter. The error bars reported reflect the standard deviation. The P values for each graph using a two-tailed analysis was determined by Graph Pad Prism version 4.0c for Macintosh.

3.5 Protein Expression of ScFv Constructs

3.5.1 Expression of ScFv Proteins in Yeast

A colony expressing a single scFv protein was grown up for 24 hours shaking in liquid SD W^- media at 30 °C. The cells were collected at 3220 x g for five minutes at room temperature and washed in 1 mL of ddH_2O. The cells were induced in 10 mL of SG/R W^- at 30 °C for eight hours and collected by centrifugation at 3220 x g.

3.5.1.1 Protein L purification of scFvs

After induction of scFv proteins and collection of the cells, the pellets were resuspended in 300 μL of protein L binding buffer (20 mM Na_2HPO_4, 0.15 M NaCl, pH 7.0), 3 μL 1 mM PMSF (Sigma Aldrich Chemical Co., Canada), and 0.3 g acid-washed glass beads (425-600 μm) (Sigma Aldrich Chemical Co.). The cells were ruptured using Fast PrepFP120 (Thermo Savant, USA) for twenty seconds and setting 6.0 at 4 °C three times. The cell lysate was centrifuged twice at 4 °C at 18 000 x g for fifteen minutes and the supernatant was applied to protein L resin (Gene Script Corporation, USA). The cell lysate was incubated on the protein L resin overnight at 4 °C gently shaking. The protein L beads and lysate were applied to a microcron centrifugal filter column (Millipore, USA) and centrifuged at 1000 x g for 30 seconds at room temperature. The beads were washed three times with binding buffer and eluted in 0.1 M citric acid pH 3.0. At each step the samples were collected and added to 2x SDS-loading buffer (62.5 mM Tri-HCl, pH 6.8, 0.25% (v/v) glycerol, 2% (w/v) SDS, 0.01% (w/v) bromophenol blue (Fischer Biotech, USA), 0.05% (v/v) β-mercaptoethanol), heated to 95 °C for 5 minutes and analyzed by Western blot

3.5.1.2 Western blotting

Cells expressing scFv proteins that were not purified by protein L beads were extracted using a protocol previously described (79). The extracted proteins were loaded into a 4% stacking gel 15% resolving gel SDS-PAGE and separated for 15 minutes at 100 V and 35 minutes at 200 V. The gel and nitrocellulose were incubated in transblot buffer, pH 8.3 (48 mM Tris, 39 mM glycine, 20% (v/v) methanol, 0.04% (w/v) SDS) for 10 minutes. The gel was electro-blotted with a transblot SD semi dry transfer (BioRad, Canada) to nitrocellulose (BioRad) at 15 V for 35 minutes. The blot was incubated in blocking buffer (Licor Biosciences, USA) for one hour and incubated with primary antibody cocktail (10 mL blocking buffer, 0.0005% (v/v) Tween 20, 60 uL HA antibody (Santa Cruz Biotechnology, USA)) overnight. The next day the blot was washed with 1 X PBS-Tween, pH 7.4 (1.37 M NaCl, 26.8 mM KCl, 10 mM $Na_2HPO_4 \cdot 7H_2O$, 17.6 mM KH_2PO_4, 0.001% (v/v) Tween 20) three times and incubated with secondary antibody cocktail (10 mL blocking buffer, 0.0005% (v/v) Tween 20, 2.5 uL anti-mouse antibody (Licor Biosciences)) for one hour. The blot was analyzed on a Licor Odessey infrared imaging system (Licor Biosciences).

3.5.2 Expression of ScFv Proteins in *E. coli*

ScFv proteins were expressed in *E. coli* to obtain more protein to analyze the scFvs.

3.5.2.1 Cloning scFvs into an inducible *E. coli* expression vector

The scFv was cloned from the yeast expression vector (pIL500) and into a bacterial expression vector (pET28b). The scFv DNA was amplified in the following reaction: 1 X PCR

Buffer, 0.2 mM dNTPs, 1 μL of pIL500 expressing a scFv, 0.4 μM P1 Intein/*Eco*RI+1 (for cyclic scFvs or P1 pET28b/VH3-74 for linear scFvs) (#41 and #39 Table 3-1), 0.4 μM P2 Intein-Npu/*Xho*I (for cyclic scFvs or P2 L19/pET28b for linear scFvs) (#42 and 40 Table 3-1), and 1 μL Taq polymerase. The PCR reaction was amplified in a thermocycler under the following conditions: 95 °C for 2 minutes, [95 °C for 30 seconds, 55 °C for 30 seconds, 72 °C for 2 minute] (25 cycles), and 72 °C for 7 minutes. pET28b and the amplified scFv PCR were digested with 0.5 units of *Eco*RI and 0.5 units of *Xho*I in a 100 μL reaction. Digested pET28b and digested scFv PCR fragments were purified by gel electrophoresis to get rid of contaminating DNA using QIAquick gel extraction kit according to manufacturer's instructions. The gel purified plasmid DNA (or vector DNA) and gel purified scFv PCR (or insert DNA) were mixed in a 1:3 molar ratio of insert DNA to vector DNA along with 1X ligation buffer (New England Biolabs) and 1 μL T4 DNA ligase (New England Biolabs) in a 20 μL reaction. The reaction was incubated at room temperature for 24 hours. Three microlitres of the ligation reaction was transformed by electroporation (80) into 50 μL of XL1Blue electrocompetent cells to obtain high quality plasmid DNA and optimal propagation of plasmid DNA. The transformed cells were grown in 500 μL of LB for one hour at 37 °C, plated on LB-Kanamycin plates and incubated at 37 °C overnight. A colony was selected from the plate and grown up in LB-Kanamycin liquid medium at 37 °C overnight. The plasmid was isolated using QIAPrep Spin Miniprep Kit according to manufacturer's instructions.

The isolated plasmid was transformed by electroporation into BL21 codon plus electrocompetent cells to allow for protein induction. The transformed cells were grown up in LB-Kanamycin liquid medium at 37°C overnight. The cells were diluted to an optical density at 600 nm of 0.2 and grown for approximately 2 hours until an optical density at 600 nm of 0.6-0.9

was reached. The cells were induced with 1 mM IPTG and grown at room temperature shaking for 24 hours. The cells were collected by centrifugation at room temperature at 3220 x g for 10 minutes.

The cells were resuspended in 600 µL of Buffer B pH 8.0 (100 mM $NaH_2PO_4 \cdot H_2O$, 10 mM Tri-Cl, 8 M urea), 0.3 g of acid-washed glass beads (425 – 600 µm) and 1 mM PMSF. The cells were ruptured using Fast PrepFP120 (Thermo Savant) for twenty seconds and 6.0 setting at 4 °C three times. The cell lysate was centrifuged at 18 000 x g for 15 minutes two times at 4 °C. The supernatant was then applied to a histidine tag-Ni Column to purify the scFv.

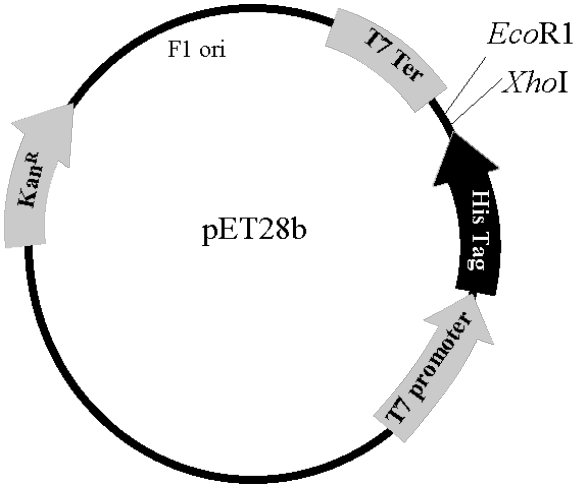


Figure 3-13: pET28b *E. coli* expression vector.
pET28b contains a T7 promoter, a T7 terminator and a F1 origin of replication (F1 ori). The scFvs cloned into this vector were cloned in through the *Eco*RI and *Xho*I restriction sites. They are expressed tethered to a poly-histidine tag (His Tag). The histidine tag consists of six consecutive histidines that can be purified with a nickel column due to the histidine tag's affinity for Nickel. It also contains a gene for kanamycin resistance allowing cells containing this plasmid to grow in media containing kanamycin.

3.5.2.2 Histidine tag-Ni column purification

The scFv was purified using the histidine tag to test the difference in stability and solubility between the linear and cyclic versions of the scFv. The extracted protein lysate containing the histidine-tagged scFv was applied to His-Select Affinity Nickel Gel (Sigma Aldrich Chemical Co.) for 30 minutes at room temperature with slow, gentle back and forth rocking. The gel and lysate were centrifuged at 1000 x g at room temperature for 30 seconds, the flow through lysate was collected. The beads were washed three times with Buffer B and eluted in Buffer B containing 500 mM imidazole (protein elution 'A'). At each step the samples were collected and separated on a 4% stacking gel 15% resolving SDS-PAGE. Protein elution 'A' was eluted through a protein L column subsequently to examine the scFv folding correctness.

3.5.2.3 Renaturation of the scFv protein

To test the correct folding of the scFvs protein elution 'A' was applied to a protein L column, which will bind scFvs if they are correctly folded. To bind to protein L the scFvs must be in their native form. The scFvs were successfully purified only under denaturing conditions so to test the correct folding they were renatured. The extracted protein lysate containing the histidine-tagged scFv were applied to His-Select Affinity Nickel for 30 minutes at room temperature with slow, gentle back and forth rocking. The beads and lysate was centrifuged at 1000 x g at room temperature for 30 seconds, the flow through lysate was collected. The beads were washed with decreasing concentrations at a 2 M gradient of urea for 30 minutes at each concentration and without urea three times. The scFvs were eluted in Buffer B containing 500 mM imidazole. At each step the samples were collected and separated on a 4% stacking gel 15% resolving SDS-PAGE.

3.5.2.4 Protein L purification of scFvs

After renaturing the scFv, the elution was applied to protein L resin and incubated overnight at 4 ºC with slow, gentle back and forth rocking. The next day the flow through was collected. The beads were washed three times with Buffer B and eluted in 0.1 M citric acid, pH 3.0. At each step the samples were collected and separated in a 4% stacking gel 15% resolving SDS-PAGE.

3.5.2.5 SDS-PAGE

Reduced and non-reduced samples of the cell lysates, washes and elutions from the beads were analyzed by SDS-PAGE. The samples were mixed with 1x SDS loading buffer, heated at 95 ºC for 5 minutes and loaded a 4% stacking gel 15% resolving SDS-PAGE and separated for 15 minutes at 100 V and 35 minutes at 200 V. The gel was incubated in fixing solution (25% (v/v) isopropanol, 10% (v/v) acetic acid) for fifteen minutes and stained in Coomasie blue stain (10% (v/v) acetic acid, 0.012% (w/v) bromophenol blue) for 4 hours. The gel was incubated in destain solution (10% (v/v) acetic acid) until the bands became clearly visible. The gel was analyzed on a Licor Odessey infrared imaging.

Chapter 4 Results

4.1 Construction of Libraries

Single chain variable fragments (scFvs) have six complementarity determining regions (CDRs), three on the heavy chain and three on the light chain (Figure 4-1). These CDRs are responsible for the highly specific interactions between the antibody and the antigen. The goal of this project is to create synthetic scFv libraries that are capable of binding a variety of antigens with high affinity and specificity. To do this, scFv libraries were constructed that varied in (i) amino acid diversity within the CDRs; (ii) the number of CDRs varied; and (iii) configuration (cyclic versus linear). The scFv libraries were produced in *S. cerevisiae*. This model organism provides a eukaryotic cell environment similar to that of humans, which is the projected destination for scFv expression. The reducing environment inside the cell does not support the disulphide bonds that stabilize the scFv structure and therefore the scFvs are not stably expressed. By expressing scFvs in *S. cerevisiae* the yeast two-hybrid assay can be used to isolate scFvs. By using this assay the scFvs isolated should be properly folded and expressed within the yeast cell.

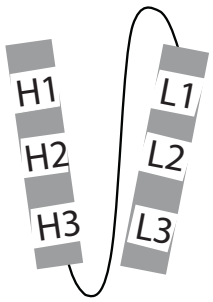
The scFvs produced in this study were based on scFvs previously reported. The scFv framework used in this study was modeled after the consensus framework found by Tse *et al*. who established an scFv framework consensus sequence that is highly expressed in yeast (74). Tanaka *et al*. showed that this framework can be used with other CDRs (73). The CDR residues varied within the framework were modeled after observations by Fellouse *et al*; who determined

important CDR residues involved in binding by solving crystal structures of antigen-antibody fragment complexes (8).

There are approximately 20 variable positions within the CDRs that come into contact with the antigen (30). A library containing scFvs with all the possible combinations of 20 amino acids in 20 different positions; $20^{20} \sim 1.0 \times 10^{26}$ far exceed the number of cells that can be produced and screened in yeast. By reducing the number of amino acids used and the variable number of CDRs, a simpler library can be produced. The goal was to maximize binding of scFvs to a variety of antigens while minimizing the size of the scFv library. To do this, scFv libraries were created based on previous results. Four or two amino acids were used to diversify the CDRs instead of the twenty found in the natural repertoire. Three or four CDRs were diversified instead of six.

Eight scFv libraries were constructed that differed in the number of variable amino acids, the number of CDRs varied and scFv configuration. The scFv libraries are listed in **Figure 4-1**. The scFv configuration within the libraries was either cyclic or linear (c or L). The letters K and T were used to represent the number of amino acids encoded by the degenerate codon, KMT or TMT. KMT encodes four different amino acids (tyrosine, serine, alanine and aspartate) and TMT encodes two different amino acids (tyrosine and serine). Therefore, in the library nomenclature T = two amino acids and K = four amino acids. There were three or four variable CDRs within each scFv library, which is represented as the number at the end of the library name. Initially, six scFv libraries (LT3, LK3, LT4, cT4, LK4, cK4) were created. These libraries were screened using the yeast two-hybrid assay to determine the effect of the variable parameters on the library binding capacity. The cyclic three variable CDR libraries were not initially created as it was assumed that the four variable libraries would have a larger binding capacity. The six libraries

were screened in three separate assays. These three screens revealed that the cyclic scFv libraries had a larger binding capacity than the linear scFv libraries. To confirm that this trend also held for the three variable CDR libraries, two additional cyclic libraries (cT3, cK3) were created.



Library Name	Configuration	Amino Acids	CDRs Varied
LT3	Linear	Tyrosine, Serine	H1, H2, H3
cT3	Cyclic	Tyrosine, Serine	H1, H2, H3
LT4	Linear	Tyrosine, Serine	H1, H2, H3, L3
cT4	Cyclic	Tyrosine, Serine	H1, H2, H3, L3
LK3	Linear	Tyrosine, Serine, Alanine, Aspartate	H1, H2, H3
cK3	Cyclic	Tyrosine, Serine, Alanine, Aspartate	H1, H2, H3
LK4	Linear	Tyrosine, Serine, Alanine, Aspartate	H1, H2, H3, L3
cK4	Cyclic	Tyrosine, Serine, Alanine, Aspartate	H1, H2, H3, L3

Figure 4-1: ScFv libraries created.
The scFv structure is shown in the left panel. The scFv is made of the variable heavy chain domain (VH) and the variable light chain domain (VL) of a natural antibody. There are three complementarity determining regions (CDRs) located on each the VH and VL. In the right panel the names of the different libraries are given in the first column. The first letter indicates linear (L) or cyclic (c), the second letter indicates number of amino acids K (four) or T (two) and the number indicates the number of CDRs.

The first parameter tested was the number of amino acids used to vary the CDRs. This parameter affects the library size immensely. For example using two or four amino acids dramatically decreases the size of the library compared to using all 20; a two amino acid library varied at twenty positions contains 2^{20} members, a four amino acid library varied at twenty positions contains 4^{20} members and a twenty amino acid library varied at twenty positions contains 20^{20}members. Studies have shown that the number of amino acids can be reduced to two amino acids (tyrosine and serine) and high affinity Fabs can still be isolated (81). In this study, the CDRs were varied using either two amino acids (tyrosine and serine) or four amino acids (tyrosine, serine, alanine and aspartate). Fellouse *et al* previously showed that this minimal

diversity is enough to create highly specific Fab binders to a variety of antigens (8,9). The number of variable CDRs also affects the library size.

Each CDR contains a different number of amino acids that contribute to antigen binding. Based on previous results, CDR L1 and CDR L2 were not varied in this study. In four of the scFv libraries, only the three heavy chain CDRs were varied and in the other four the heavy chain CDRs as well as CDR L3 was varied (Figure 4-1). The residues within the CDRs were varied based on the scFv used by Fellouse *et al* (Figure 3-5) (8). By decreasing both the number of CDRs varied and the amino acids varied within them, the diversity of the library was greatly reduced.

The configuration does not affect library diversity or size; it is proposed to affect the stability of the library and, therefore is proposed to increase library functionality by increasing the number of properly folded scFvs. The functional diversity is described as the approximate number of the scFvs within an scFv library that contains the correct sequence information. The functional diversity was determined as a percentage of scFvs with the correct sequence out of a sample of 50 scFvs that were sequenced from each library. The scFv structure is inherently unstable since it is only stabilized by hydrophobic interactions between the two variable domains. The intra-domain disulphide bonds, which normally assist in stability, do not form within the cell (Figure 4-1) (37). Thus intracellular scFvs can unfold leading to aggregation and decreased solubility, which hinders binding (41). By cyclizing the scFv it should prevent aggregation and increased stability (12,13). This should increase the number of properly folded scFvs within the library and the ability of the scFvs to interact with their target; thus, increasing the binding capacity of the cyclic libraries versus the linear libraries.

The libraries were constructed using yeast homologous gap repair mechanism in a series of steps. The scFv library framework with restriction sites in place of the heavy chain CDRs was constructed from eighteen overlapping oligonucleotides by PCR amplification. CDR H3 was PCR amplified and cloned into the scFv framework. This plasmid was amplified and CDR H1 and H2 were cloned into the scFv framework containing CDR H3. This gave rise to the linear scFv libraries containing three variable CDRs (K3, T3). The four variable CDR scFv libraries were made from the three variable CDR scFv libraries. The three variable CDR scFv libraries were PCR amplified using mutagenic primers that diversify CDR L3. The primers contained degenerate codons to add diversity to CDR L3. The PCR amplified scFv libraries were cloned into the plasmid (pIL500) generating the linear four variable CDR libraries (K4, T4). The cyclic libraries were created from the linear libraries. To create the linear libraries were PCR amplified and cloned into a vector containing the intein machinery necessary for protein cyclization. This created a slight artificial bias against the libraries containing four variable CDRs and the cyclic libraries that were PCR amplified from the libraries with three variable CDRs and linear libraries, which potentially reduces their diversity.

4.1.1 Diversity of libraries

The theoretical diversity of each library is indicated in Table 4-1 and is calculated by taking the number of varied amino acids raised to the power of the total number of variable residues within the CDRs. The obtained diversity of the scFv libraries is determined by the efficiency of the yeast transformation. The scFv libraries were created using three different cloning steps. The transformation efficiency at the last step was recorded as the library diversity. It is assumed every successful transformant is a unique scFv library member containing a

different CDR sequence. Fifty library members were sequenced from each library to analyze the library diversities. All scFvs analyzed had different CDR sequences. From the fifty sequences obtained from each of the libraries, a percent of functional library members was calculated. The functional percentage was multiplied by the obtained diversity to obtain the functional diversity. The functional diversity represents the number of scFvs within the library with the correct sequence.

Table 4-1: ScFv library diversities.

Library	Variable CDRs	Vaiable amino acids	Theoretical Diversity	Obtained Diversity	Functional (%)	Functional Diversity
LK3	18	4	6.9×10^{10}	5.4×10^{7}	73	3.9×10^{7}
cK3	18	4	6.9×10^{10}	2.0×10^{7}	60	1.2×10^{7}
LK4	23	4	7.0×10^{13}	2.0×10^{7}	52	1.0×10^{7}
cK4	23	4	7.0×10^{13}	2.2×10^{6}	41	9.0×10^{5}
LT3	18	2	2.6×10^{5}	8.0×10^{5}	67	5.4×10^{5}
cT3	18	2	2.6×10^{5}	7.0×10^{5}	60	4.2×10^{5}
LT4	23	2	8.4×10^{6}	4.2×10^{6}	51	2.1×10^{6}
cT4	23	2	8.4×10^{6}	4.2×10^{6}	65	2.7×10^{6}

*The name of the library is indicated in the first column where L = linear, c = cyclic, K = four amino acids, T = two amino acids, 3 = 3 CDRs, and 4 = 4 CDRs. The variable number of amino acid residues within complementarity determining regions (CDRs) and the number of amino acids varied in those positions are indicated in the second and third columns. The theoretical diversity is the total number of possible unique members. The theoretical diversity was calculated using the number of amino acids varied raised to the power of the number of varied CDRs. (ie LK3, 23 different positions, 4 amino acids; $4^{23}= 7 \times 10^{13}$). The obtained diversity is the number of transformants obtained after the final transformation. The functional diversity is an approximate percent of scFvs with the correct scFv sequence taken from a sample of fifty members from each library. The functional diversity is the product of the obtained diversity and the functional percent to give the approximate number of unique and functional scFvs from each library.

The functional percentage is used to normalize the number of functional scFvs mated and screened. For example, in the LK3 library, if 1×10^{6} cells from each library are screened, the

number of LK3 cells that need to be plated in order to have 1 x 10^6 functional scFvs is [(1 x 10^6 x 0.27) + 1 x 10^6] = 1.27 x 10^6. All the library screens are adjusted in this manner.

4.1.2 Cyclization of scFv libraries

ScFvs are inherently unstable inside cells. ScFvs contain disulphide bonds that stabilize their structure but they are not formed inside of cells. The limited intracellular stability of scFvs is due to the hydrophobic bonds between the two variable domains (37). Without the extra stability from the intra-domain disulphide bonds, scFvs have a tendency to unfold and aggregate. When aggregated, scFvs precipitate out of solution and are not functional. Studies have shown that cyclization of proteins can increase stability, solubility, and/or folding ability while still retaining activity (12,13). The intein is a self-splicing protein whose potential has been harnessed to successfully cyclize proteins *in vivo* (Figure 2-5) (54).

In attempt to stabilize the scFv structure, linear scFv libraries were cloned into a vector containing the intein machinery that has been used to successfully cyclize proteins in previous reports (58) and in our lab. ESI-TOF mass spectroscopy was used to analyze the cyclization of single scFvs. The analysis of the scFvs by mass spectrometery was not successful, which could be due to the size of the scFv or the scFv not being appropriately charged. To see if the cyclic and linear scFv were being expressed and the intein was processing, they were separated by SDS-PAGE (Figure 4-2). The two scFvs were identical in their CDR sequences and only differed in configuration. They were denatured, purified using nickel beads and separated by SDS-PAGE.

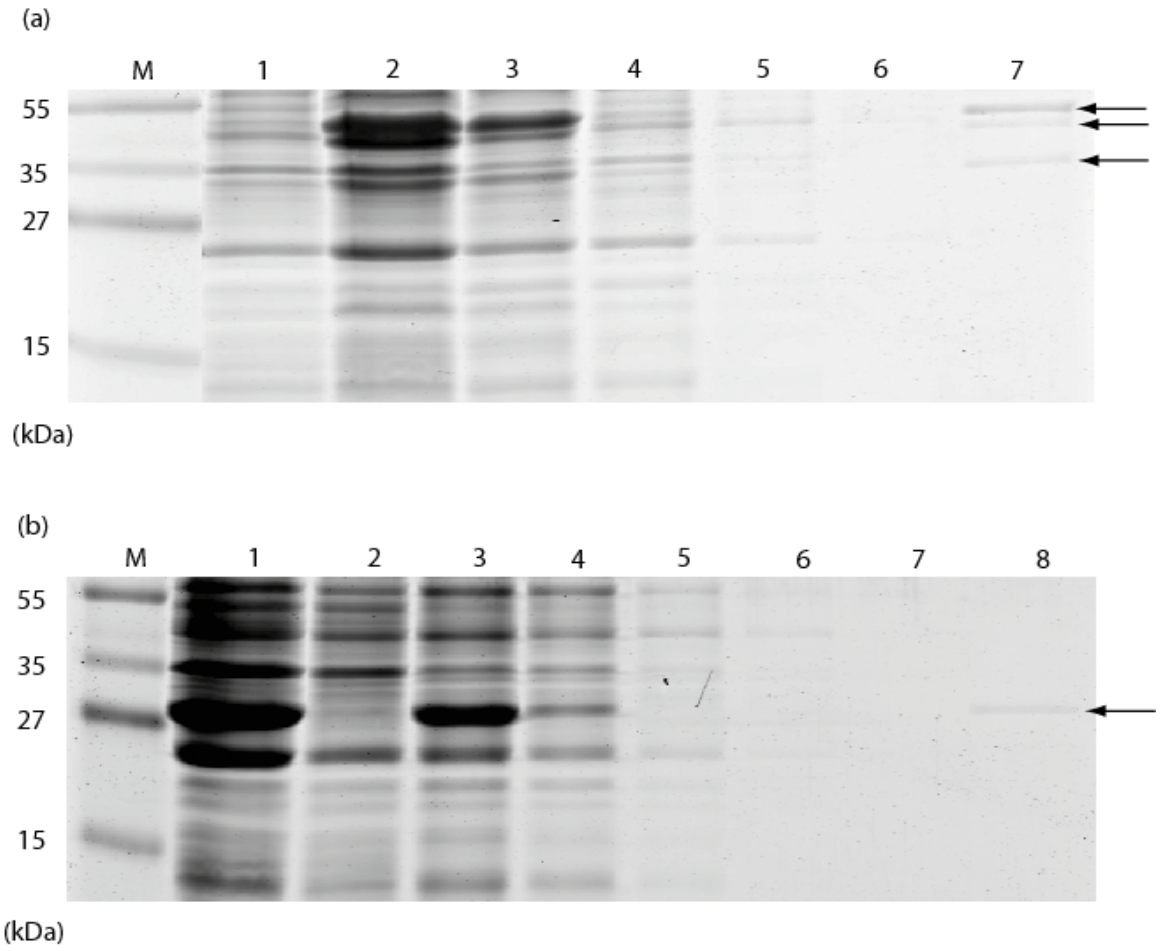


Figure 4-2: SDS-PAGE analysis of scFvs expressed in *E. coli*.
A cyclic and linear scFv, containing the same CDRs were induced for 24 hours and bound to nickel beads. The beads were washed, eluted and separated in a 4% stacking 15% resolving SDS-PAGE and stained with coomassie blue stain. (a) Cyclic scFv; lane M, standard pre-stained protein marker; lane 1, uninduced; lane 2, induced; lane 3, crude cell lysate before incubation with nickel beads; lane 4, flow through lysate that did not bind the nickel beads; lane 5, wash 1; lane 6, wash 2; lane 7, elution from the nickel beads. (b) Linear scFv; lane M, standard pre-stained protein marker; lane 1, induced; lane 2, uninduced; lane 3, crude cell lysate before incubation with nickel beads; lane 4, flow through lysate that did not bind the nickel beads; lane 5, wash 1; lane 6, wash 2; lane 7, wash 3; lane 8, elution from the nickel beads. The arrows indicate the major protein products purified by the nickel beads. In (a) the top bands is the unprocessed, the middle band may be a degraded version of the unprocessed and the lower band is the cyclized version of the scFv. In (b) the arrow indicates the linear version of the scFv.

The linear and cyclic scFvs are approximately 31 kDa and 36 kDa, respectively. There were three major bands in the cyclic scFv elution, one that ran around 35 kDa and the other two ran in between the 35 and 55 kDa markers (Figure 4-2). The cyclic scFv is 36 kDa, which corresponds with the lower band. The unprocessed scFv (scFv + N-intein + C-intein) is approximately 48 kDa. Depending on cyclization efficiency there will usually be some unprocessed scFv present. The largest band in the cyclic gel corresponds to the unprocessed scFv and the other band is likely the folded unprocessed or a splicing intermediate of the unprocessed intein. The third band in the elution corresponds to the cyclic scFv. There is one major band in the linear gel that ran at approximately 31 kDa, which corresponds with the linear scFv. Since the unprocessed and processed versions of the scFv are visible, it indicates that the intein is processing the scFv, which is a good indication that the scFv is being cyclized.

4.2 Library Screening

To evaluate the scFv libraries for binding capacity, several protein domains were chosen to screen as targets against each of the libraries using the yeast two-hybrid assay (Figure 2-6). Five different screens were done with different protein domains and the libraries at different mating densities (Table 3-4). To screen the libraries, yeast cells expressing the protein domains and yeast cells expressing the library were separately grown to OD_{600} = 0.6-0.9. The bait cells were pooled and mated separately with each library. The mated cells were screened first on weak selection media (SG/R $H^-W^-L^-$) and later replica plated to a stronger selection media (SG/R $H^-W^-A^-$ -X-gal). This two-step screening procedure has been shown previously to reduce false positive results in the yeast two-hybrid assay (83).

4.2.1 Protein domains used in the interaction mating assay

The protein domains used in the interaction mating assay portion of this study are domains from BCR-ABL and HCK (Table 3-4). Both BCR-ABL and HCK contribute to chronic myelogenous leukemia (CML). BCR-ABL is an oncogene fusion product that is found in CML signaling pathways (84). BCR-ABL is created by a reciprocal translocation between chromosomes 9 and 22 to produce the novel fusion protein BCR-ABL (84). HCK is one of the proteins that BCR-ABL activates and sends signals through. HCK is activated by BCR-ABL, which in turn phosphorylates BCR-ABL and other targets spreading the transforming signal (85). HCK and BCR-ABL have been shown to interact through the SH3 domain of HCK and SH1, SH2, and SH3 domains of BCR-ABL, as well as the C-terminal domain of ABL which contains the F-actin binding domain (86). BCR-ABL and HCK were used as targets in the yeast two-hybrid assay to evaluate the scFv libraries.

Five out of the seven domains were used in each screen. The domains used from BCR-ABL included: the coiled coil domain (cc1-72), src homology 2 (SH2), src homology 3 (SH3), poly-proline (PxxP), tyrosine residue 177 flanked by ten amino acid residues (Y177), and f-actin binding domain (FABD). The SH1 tyrosine kinase domain from HCK was also used. HCK and BCR-ABL were split into domains because they are very large proteins and targeting an single protein domain leads to a more specific protein interaction, which will allow the user to better predict the function of the targeting protein in a cell. For example, since BCR-ABL and HCK interact through BCR-ABL SH1, SH2 and SH3, obtaining an scFv specific for one of these domains should have an effect on this interaction, whereas blindly targeting the whole protein does not ensure that an important interaction will be disrupted. The domains used in this study are structurally diverse to prevent any bias among the binding capabilities of the different

libraries. The cc1-72 domain is made up of two or three α-helices (87). The SH2 is made up of three-stranded anti-parallel β-sheets and two flanking α-helices (88,89). The SH3 domain consists of five β-strands that form a β-sandwich with two tightly packed anti-parallel β-sheets with one long β-hair-pin turn (89). Y177 is a tyrosine that is involved in downstream signaling (90,91). The Y177 domain comprises of eleven amino acids in total with tyrosine 177 in the middle. The PxxP is a poly-proline rich region (92). The FABD is made up a 4 α-helical bundle (93). These different protein domains were used as baits to screen the scFv libraries.

4.2.2 Screen Results

The output of the yeast two-hybrid assay measures the 'binding capacity' or number of interactions per number of diploids screened. It is difficult to test the binding capacity of the libraries because of the large range in diversity between the libraries. To address this issue, the libraries were screened five times, each with a different amount of library members (Table 3-4). Initially, six scFv libraries (T3, T4, cT4, K3, K4, cK4) were made to test the differences imposed by the different parameters (Figure 4-1). Since the smallest library contained 2.6×10^5 members, the first screen was done with 2.6×10^5 library members.

The six libraries were screened two more times with one million and six million library members to compare the results obtained in the first screen (Table 3-4). There was no clear trend with respect to number of amino acids or number of variable CDRs; however, the first three screens showed in all cases that the cyclic scFv libraries contained a larger binding capacity than the linear.

The cT3 and cK3 libraries were made after these compelling results regarding scFv cyclization. In the fourth screen twelve million library members from each library were screened

which included the cT3 library. The cyclic scFv libraries again contained a higher binding capacity compared to the linear libraries. The cK3 library was produced and screened along with the seven other libraries using twenty million scFv library members in the fifth screen. The trend observed among the first four screens persisted, showing that the cyclized libraries had a larger binding capacity than the linear libraries. This reoccurring trend was of particular interest as it alluded to the fact that cyclization was stabilizing the scFv structure enabling them to bind more targets.

4.2.3 Cyclic versus linear in binding capacity

When expressed inside a cell, the linear scFv structure is only stabilized by hydrophobic interactions between the two domains and the linker peptide, which tethers the two domains together (5). For this reason scFvs are unstable, prone to unfolding, and form aggregates, which makes most scFvs insoluble (39,40). When proteins form aggregates, they are unable to function because they precipitate out of solution (39,40).

In this study, four cyclic scFv libraries and their linear counterparts were created to determine what effect cyclizing has on their binding capacity. The cyclic scFv libraries had a higher binding capacity than the linear libraries in all five screens (Figure 4-3). To compare the different screens, each screen was normalized by taking the ratio of the cyclic library over the linear library. In all five screens, the cyclic scFv libraries have a statistically significant difference (P values < 0.05) in binding capacity when compared to the linear scFv library counterpart. Not only was the difference between the cyclic and linear library in each screen significant the cyclic libraries always had a higher binding capacity than the linear libraries.

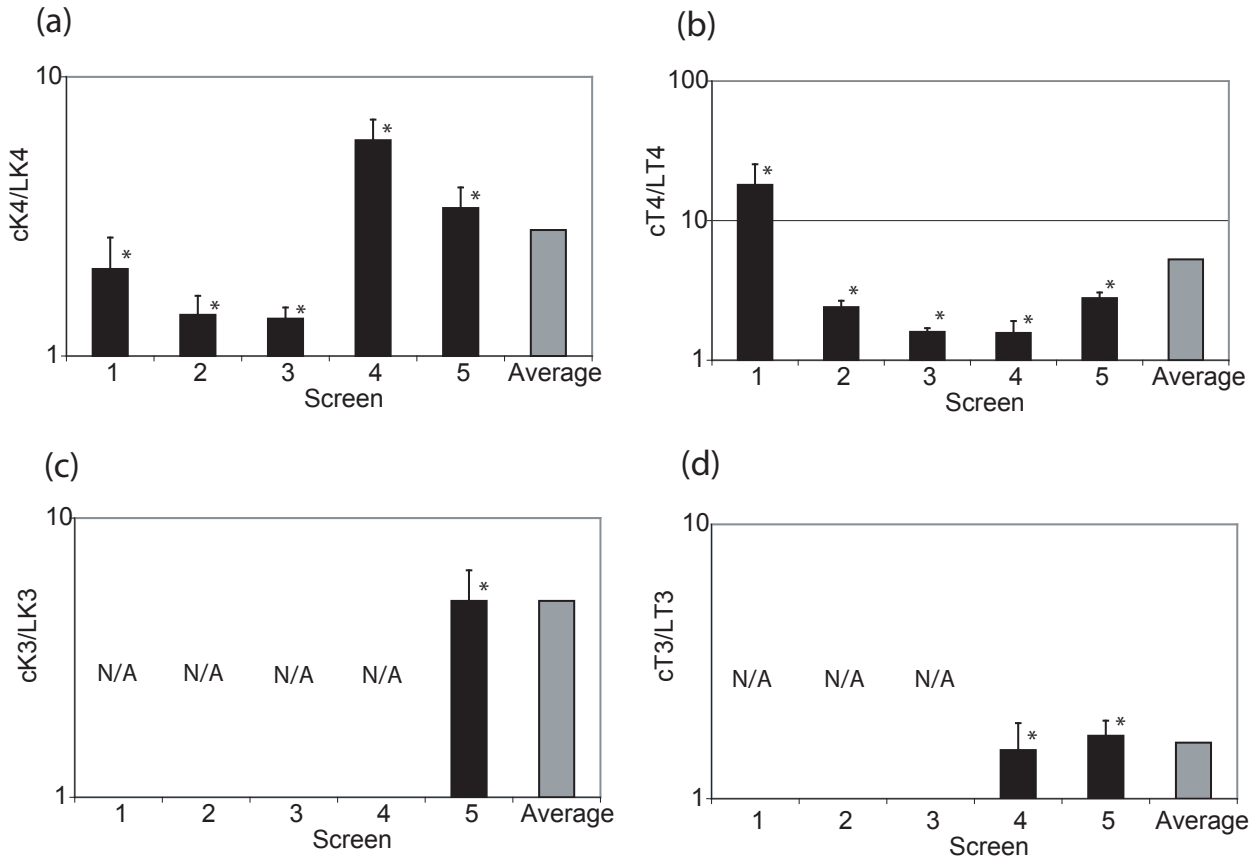


Figure 4-3: Ratios comparing cyclic configuration to linear**.**
The screens are listed along the x-axis; screen 1 = 2.6 x 10^5 library members, 2 = 1 x 10^6 library members, 3 = 6 x 10^6 library members, 4 = 12 x 10^6 library members, and 5 = 20 x 10^6 library members. The ratio of positive interactions per diploid from each screen is represented on the y-axis as the cyclic interactions divided by the linear interactions. The error bars represent the standard deviation from five plates used for each screen. The average of the ratios is indicated in the lane six. The asterisk (*) indicates statistical significance (P value < 0.05) between the cyclic and linear scFv library binding capacities. Data is not available (N/A) for some of the screens with cT3 and cK3, which were not yet produced at the time of the screen. (a) The cyclic four amino acid four CDR library (cK4) compared to the linear four amino acid four CDR library (LK4). (b) The cyclic two amino acid four CDR library (cT4) compared to the linear two amino acid four CDR library (LT4). (c) The cyclic four amino acid three CDR library (cK3) compared to the linear four amino acid three CDR library (LK3). (d) The cyclic two amino acid three CDR library (cT3) compared to the linear two amino acid three CDR library (LT3).

Cyclizing scFv libraries increased their binding capacity. This result is novel for scFvs and very useful for the generation of future scFv libraries. The increased binding capacity of the cyclic libraries can be attributed to the stabilization of the scFv structure.

The results found from these screens in regard to cyclic and linear are comparable because the sizes of the libraries being compared (cyclic vs. linear) are similar in size. When looking at functional diversities the linear T libraries are similar sizes to their cyclic counterpart (Table 4-1). The K libraries diversities are very different when comparing the corresponding cyclic and linear libraries. The cK4 library is less than one tenth of the size of the LK4 and the cK3 library is approximately one quarter the size of the LK3. This is interesting because despite the lower functional diversity of the cyclic libraries they still had larger binding capacities. This is strong evidence that cyclizing the scFv structure increases its folding capability and its stability, thus increasing its binding capacity. Whether cyclizing the scFv increases its stability and solubility compared to the linear version will have to be measured to confirm these results.

4.2.4 Binding capacity of three and four variable CDRs

To determine if more CDR diversity results in a larger binding three or four out of six CDRs were varied. Important CDRs within the antibody-antigen complex vary depending on the complex being analyzed. The heavy chain interacts with its target more than the light chain (31). Each CDR contains a different number of amino acids that contribute to antigen binding. For example, CDR H3 has been shown to be the most important in antigen interactions (30,31,32). CDR H3 is found in the middle of the binding pocket and is the most variable CDR in terms of size, conformation, and composition (32). CDR H2 and CDR H1 contribute more than the light chain but less than CDR H3 in binding to the antigen (30,31,32). Interactions between the light chain and antigen are also dependent on the complex being analyzed. CDR L3 contacts the antigen the most CDR L1 and CDR L2 contributes less or sometimes not at all (82).

Using these previous results as a guide, CDRs H1, H2, and H3 of the heavy chain were varied in four libraries and CDRs H1, H2, H3 from the heavy chain and L3 from the light chain were varied in four libraries (Figure 3-5). It was hypothesized that the four variable CDR libraries would have a higher binding capacity than the three variable CDR libraries because they are more diverse. The trend when comparing the binding capacity of the scFv libraries with three CDRs and the scFv libraries with four did not follow this logic. In 11 out of 13 screens, the three CDR libraries had a higher binding capacity than the four CDR libraries (Figure 4-4). Seven out of ten had a statistically significant difference (P value < 0.05). The two amino acid three CDR libraries had a larger binding capacity than the two amino acid four CDR libraries. It is difficult to comment on the two amino acid cyclic libraries when only two screens were done, however, four out of five screens with the two amino acid linear libraries showed that three CDRs were better than four, three of these were statistically significant (P value < 0.05).

The problem with fully accepting that three CDRs are better than four for the two amino acid libraries is the two cases where the four CDRs were statistically better (P value < 0.05). With the exception of these two cases these results concluded that the three CDR libraries had a larger binding capacity than the four CDR libraries.

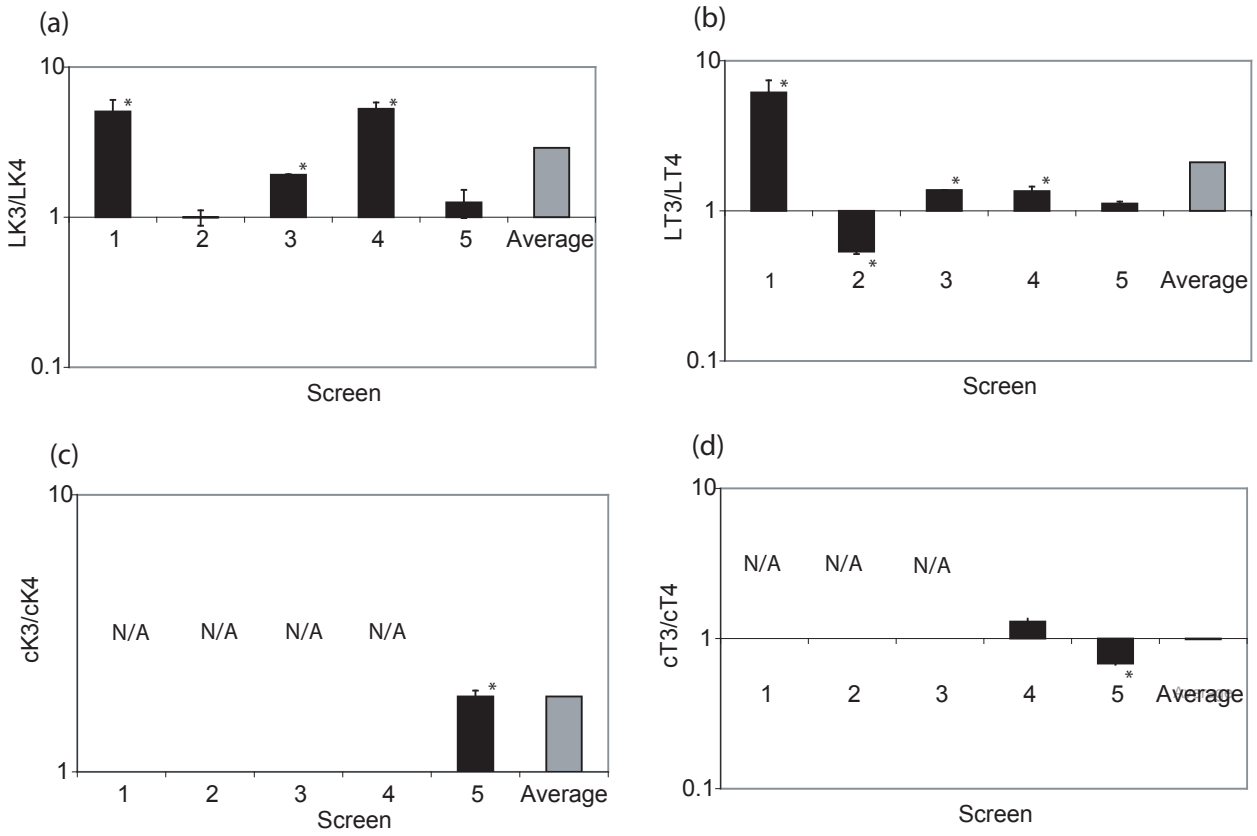


Figure 4-4: Ratios comparing three CDRs to four CDRs**.**
The screens are listed along the x-axis; screen 1 = 2.6 x 10^5 library members, 2 = 1 x 10^6 library members, 3 = 6 x 10^6 library members, 4 = 12 x 10^6 library members, and 5 = 20 x 10^6 library members. The ratio of positive interactions per diploid from each screen is represented on the y-axis as the number of interactions from the three CDR libraries divided the number of interactions from the four CDR libraries. The error bars represent the standard deviation. The asterisk (*) indicates statistical significance (P value < 0.05). The average of the ratios is indicated in the lane six. Data is not available (N/A) for some of the screens with cT3 and cK3, which were not produced at the time of the screen. (a) The linear four amino acid three CDR library (LK3) compared to the linear four amino acid four CDR library (LK4). (b) The linear two amino acid three CDR library (LT3) compared to the linear two amino acid four CDR library (LT4). (c) The cyclic four amino acid three CDR library (cK3) compared to the cyclic four amino acid four CDR library (cK4). (d) The cyclic two amino acid three CDR library (cT3) compared to the cyclic two amino acid four CDR library (LT4).

The scFv libraries with three diversified CDRs had a larger binding capacity than the libraries with four. This result could be attributed to there being less functional scFvs when the CDR L3 is diversified. For example, diversifying the fourth CDR, CDR L3 could decrease the

ability of the scFvs to fold properly and therefore decrease the number of functional scFvs in the library (compared to the scFv libraries with three diversified CDRs). Diversifying the fourth CDR negatively affected the scFv folding so much that the binding capacity was not compensated by the increased diversity from the fourth CDR of the scFvs that do fold properly.

4.2.5 Two variable amino acids versus four in CDR variability

Minimal libraries produced with only four amino acids diversifying the CDRs are able to generate high affinity antibody fragments specific for all targeted antigens (8). The largest number of scFvs was isolated from the library with CDRs diversified with tyrosine, serine, alanine, and aspartate (8). Structural analysis revealed that tyrosine side chains are responsible for mediating most antigen recognition in Fab binding pockets and small amino acids like serine and alanine allowed for space and conformational flexibility for larger amino acids like tyrosine (8). To test this, smaller libraries with only tyrosine and serine randomizing the CDRs were created (9). The serine-tyrosine library was able to isolate specific interactions against a broad range of targets.

The libraries created for this study were modeled from previous results with the goal of producing scFv libraries with decreased diversity without sacrificing binding capacity. Eight scFv libraries were created. To decrease the diversity the scFv CDRs were varied with two or four amino acids. The CDRs were varied with serine and tyrosine or serine, tyrosine, alanine, and aspartate. The hypothesis was that the four amino acid libraries would have a higher binding capacity.

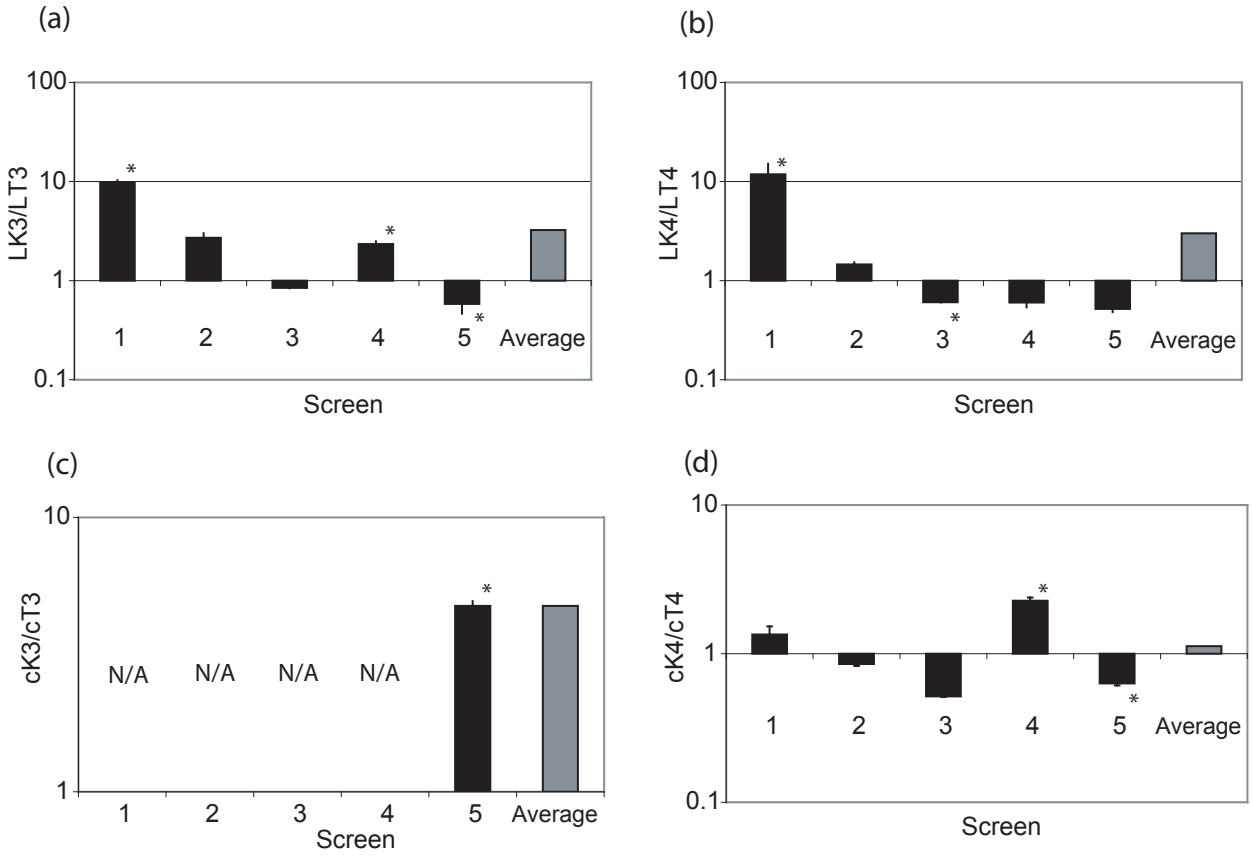


Figure 4-5: Ratios comparing two amino acids to four amino acids**.**
The screens are listed along the x-axis; screen 1 = 2.6 x 10^5 library members, 2 = 1 x 10^6 library members, 3 = 6 x 10^6 library members, 4 = 12 x 10^6 library members, and 5 = 20 x 10^6 library members. The ratio of positive interactions per diploid from each screen is represented on the y-axis as the interactions from the four amino acid libraries divided by the interactions from the two amino acid libraries. The error bars are represented by the standard deviation. The asterisk (*) indicates a statistical significance (P value < 0.05). The average of the ratios is indicated in the lane six. Data is not available (N/A) for some of the screens with cT3 and cK3, which were not yet created at the time of the screen. (a) The linear four amino acid three CDR library (LK3) compared to the linear two amino acid three CDR library (LT3). (b) The linear four amino acid four CDR library (LK4) compared to the linear two amino acid four CDR library (LT4). (c) The cyclic four amino acid three CDR library (cK3) compared to the cyclic two amino acid three CDR library (cT3). (d) The cyclic four amino acid four CDR library (cK4) compared to the cyclic two amino acid four CDR library (cT4).

The results regarding the two and four amino acid libraries showed that the libraries with three diversified CDRs, four amino acids libraries had a larger binding capacity than the two amino acid libraries (Figure 4-5). There was statistically significant differences (P value < 0.05)

supporting this observation in three out of six screens that show this trend. In the libraries with four diversified CDRs there was no difference in the binding capacity when comparing two and four amino acids. Statistically significant differences between the four CDR libraries comparing two and four amino acids were random. This is expected if neither two nor four amino acid libraries are better. When correlating these results with the other results in this study it showed that the three CDR libraries have improved binding capacity when the diversity is increased from two to four amino acids. This same trend was not observed when the fourth CDR, CDR L3 was diversified. There was no trend in the four CDR libraries when comparing two and four amino acids. This extra diversity can be attributed to the possible decrease in functional diversity (or folding of the scFvs), which is not compensated by the extra diversity from the fourth CDR.

4.3 Expression of scFv constructs

The scFvs were expressed in *E. coli* and *S. cerevisiae* to address if they are cyclizing and if they are more stable due to cyclization.

4.3.1 ScFv expression in *S. cerevisiae*

ScFvs isolated using the yeast two-hybrid assay were expressed in *S. cerevisiae*. It was hypothesized that the *S. cerevisiae* expression system would be the more likely system to produce correctly folded, active, and stable scFvs than *E. coli* because they were isolated from *S. cerevisiae*. Complex foreign eukaryotic protein expression is generally more successful in yeast compared to *E. coli*. ScFvs are difficult to express in bacteria due insolubility and formation of inclusion bodies (94). *E. coli* cells are limited in foreign protein expression because most proteins must be denatured before purification (95).

4.3.1.1 ScFv expression time course in *S. cerevisiae*

ScFvs isolated from the yeast two-hybrid screens were used to analyze expression in yeast. The scFvs are expressed in *S. cerevisiae* from a galactose inducible promoter. The scFv is expressed tethered to a hemagglutin (HA) tag and an activation domain (Figure 4-7). The cyclic scFv is also expressed with the C-intein and the N-intein on the N-terminal and C-terminal ends of the scFv respectively. Cells expressing scFvs were induced with galactose for 24 hours. A time course of the expression of a cyclic scFv was monitored to determine the optimal induction time (Figure 4-6**).** Each time point was separated by SDS-PAGE and probed with α-HA antibody, to visualize the proteins containing the HA tag.

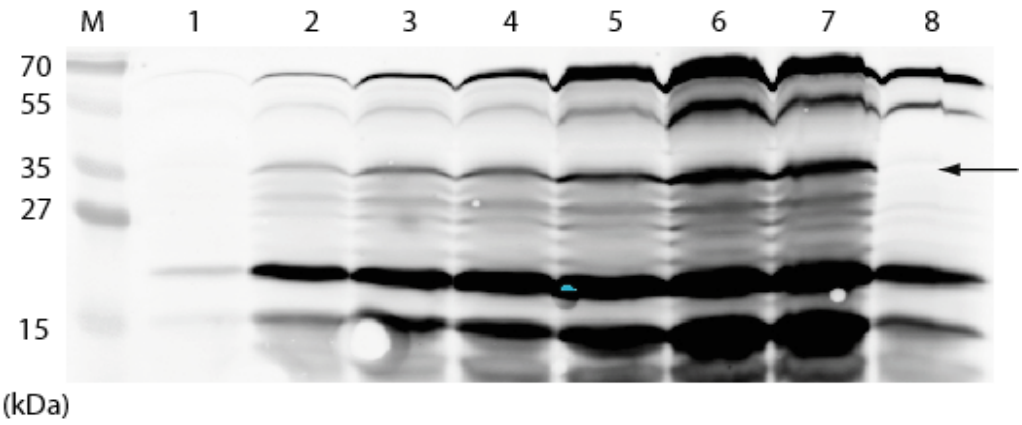


Figure 4-6: Cyclic scFv expressed in *S. cerevisiae*.
The scFv was induced for 24 hours and time points were taken. The cells were treated with NaOH rapid protein purification from yeast protocol as previously described (79). Fifteen μL of each sample was separated in a 4% stacking 15% resolving SDS-PAGE and Western blotted with α-HA antibody. Lane M, a pre-stained protein marker; lane 1, one hour of induction; lane 2, three hours of induction; lane 3, four hours of induction; lane 4, five hours of induction; lane 5, six hours of induction; lane 6, eight hours of induction; lane 7, nine and a half hours of induction; lane 8, 25 hours of induction. The arrow indicates where the cyclic scFv is.

The cyclic scFv is approximately 42 kDa. There is a band that corresponds to this size shown in Figure 4-6. The expression of the cyclic scFv was optimized at 8 hours. The

unprocessed scFv was visible at approximately 55 kDa; its expression was most intense at eight hours. There was unprocessed and cyclic scFvs present, which indicated that the intein machinery is processing to produce cyclic scFvs.

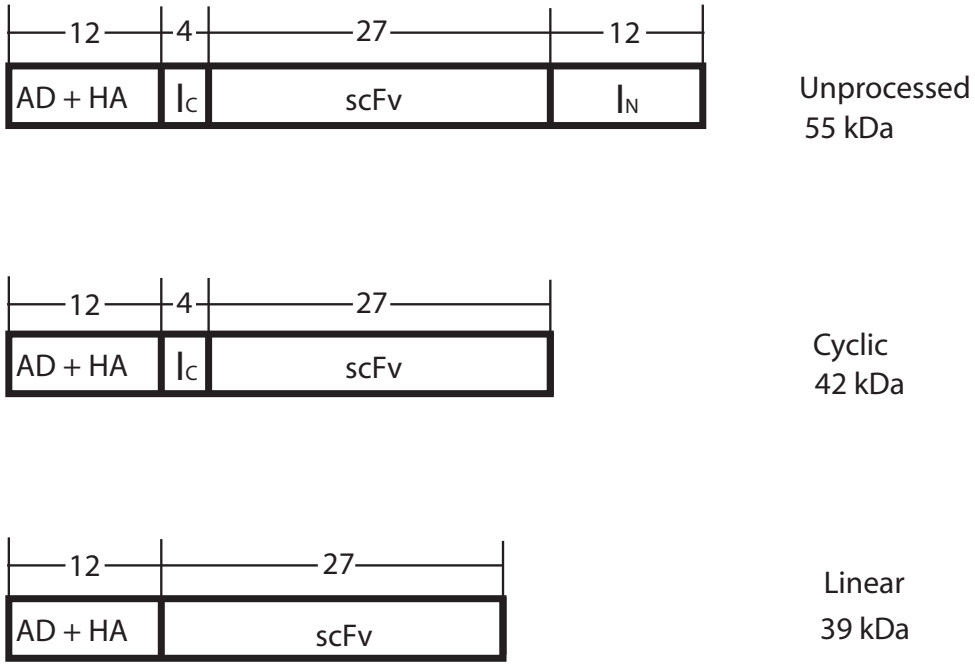


Figure 4-7: ScFv constructs expressed in *S. cerevisiae*.
The unprocessed version of the scFv-intein will be visible where the intein mediated scFv cyclization reaction has not occurred. The unprocessed version is composed of the activation domain (AD) which is part of the yeast two-hybrid machinery that activates transcription, the C-intein (I_C) and N-intein (I_N) catalyze the intein-mediated cyclization process, Hemagglutin tag (HA) for purifying and identifying the protein, and the scFv the protein of interest. The unprocessed scFv is approximately 55 kDa. The cyclic version of the scFv differs from unprocessed in that it has undergone the cyclization reaction cleaving off the N-intein domain ultimately cyclizing the scFv making the cyclic scFv approximately 42 kDa. The linear version of the scFv does not contain any of the of the intein machinery and is approximately 39 kDa.

4.3.1.2 ScFv protein L purification

Protein L is a protein isolated from *Peptostreptococcus magnus* that has a natural ability to bind to antibody kappa light chain variable regions in their native conformation (52). Protein L can be used to purify antibodies and antibody fragments, such as scFvs from solution. ScFvs isolated from the yeast two-hybrid screen were induced for eight hours with galactose. The cells were ruptured and applied to protein L beads. The resin was washed, eluted and the samples were analyzed using a Western blot (data not shown). The scFv did not bind to the protein L column. A few different conditions were used, changes in buffers, incubation times and temperatures were made but still no binding was observed.

4.3.2 ScFv expression in *E. coli*

Expression of the scFvs in *S. cerevisiae* was expected to be more successful, however purifying the scFvs using protein L from was not successful. To obtain scFvs they were expressed in *E. coli* and purified using the histidine tag. ScFvs are not usually soluble in *E. coli* but have previously been shown to have a higher expression and more efficient purification from *E. coli* than *S. cerevisiae* (94).

4.3.2.1 ScFv expression time course in *E. coli*

Some of the scFvs isolated from the yeast two-hybrid screens were cloned into *E. coli* expression vector pET28b. pET28b is a *E. coli* protein expression vector. It contains a kanamycin resistant gene for plasmid selection and a histidine tag for protein purification. The cells were grown up in LB-kanamycin media, induced with IPTG at room temperature and time points were taken. The cells were ruptured under native and denaturing conditions, separated by SDS-PAGE, and stained with coomassie blue stain (Figure 4-8).

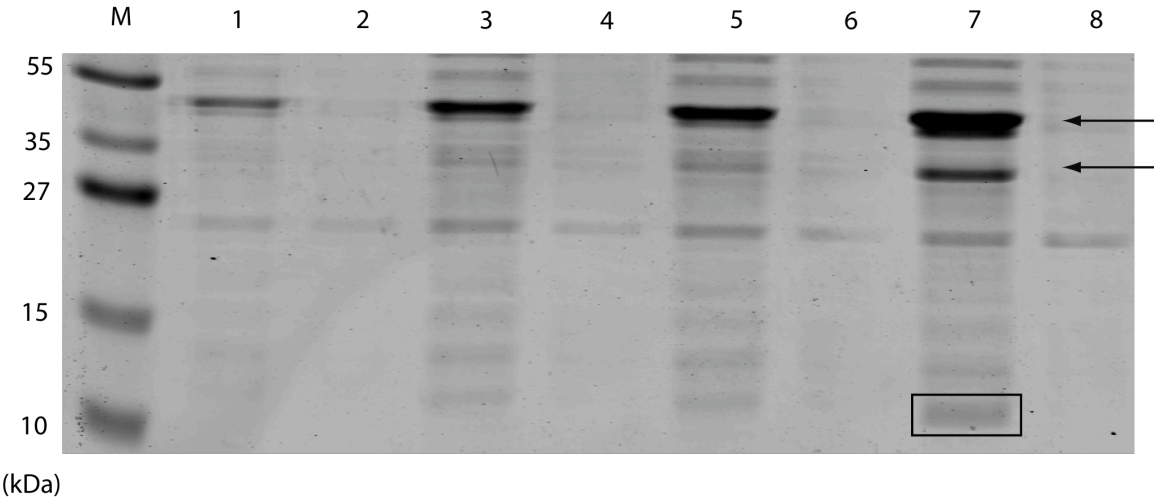


Figure 4-8: ScFv expression in *E. coli*.
The scFv was induced for 24 hours and time points were taken. The cells were ruptured under native and denaturing conditions. Fifteen µL of each sample was separated in a 4% stacking 15% resolving SDS-PAGE and stained with Coomassie blue dye. Lane M, a pre-stained protein marker; lane 1, denatured scFv at three hours of induction; lane 2, natively expressed scFv and three hours of induction; lane 3, denatured scFv at four hours of induction; lane 4, natively expressed scFv at four hours of induction; lane 5, denatured scFv at five hours of induction; lane 6, natively expressed scFv at five hours of induction; lane 7, denatured scFv at 21 hours of induction; lane 8, natively expressed scFv at 21 hours of induction. The lower arrow indicates where the cyclic scFv is approximated to run (36 kDa). The upper arrow indicates where the unprocessed cyclic scFv is approximated to run (48 kDa). The box highlights the N-intein domain.

When expressed in *E. coli*, the cyclic scFv is approximately 36 kDa. (Figure 4-9) The best induction time for the cyclic scFv was 21 hours. More protein was isolated under denaturing conditions. The cyclic scFv was not present until 21 hours. This is similar to other reports that have induced the scFv for 16 hours at 24 °C (96). The unprocessed scFv will be present depending on the efficiency of the cyclization reaction. The unprocessed scFv is visible at approximately 48 kDa. There was a band that corresponds to the size of the N-intein domain (12 kDa) (Figure 4-8). Since the N-intein domain is present it indicates that the intein is processing.

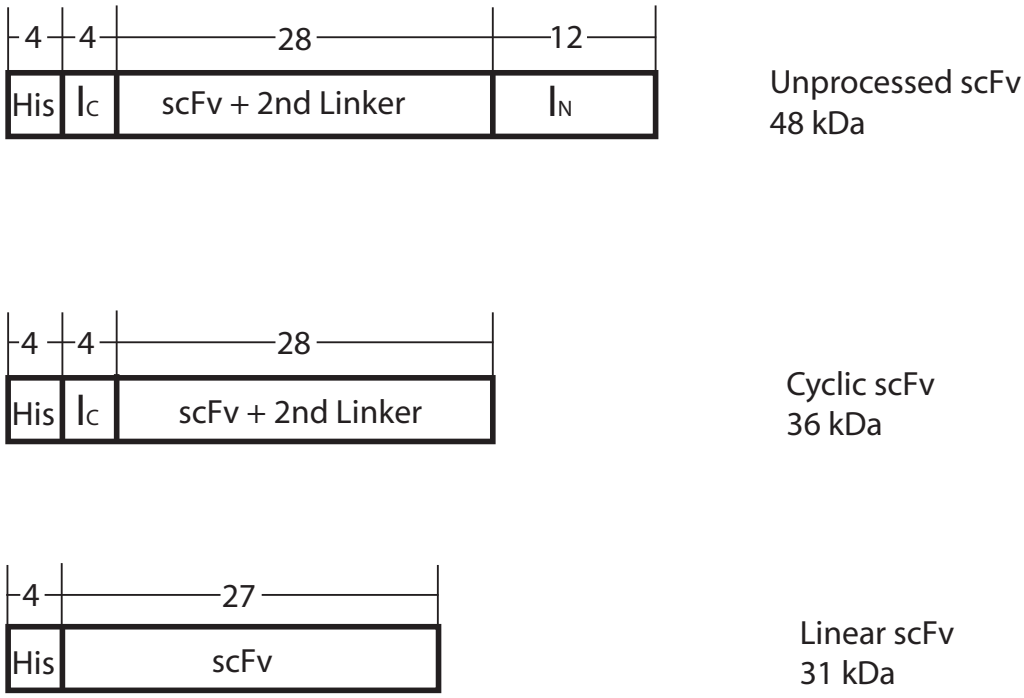


Figure 4-9: ScFv constructs expressed in *E. coli.*
The unprocessed version of the scFv-intein will be visible where the intein mediated scFv cyclization reaction has not occurred. The unprocessed version in composed of the 6x Histidine tag (His) which is used to purify and identify the scFv, the C-intein (I_C) and N-intein (I_N) used in the intein-mediated cyclization process, The unprocessed scFv is approximately 48 kDa. The cyclic version of the scFv differs from unprocessed in that it has undergone the cyclization reaction cleaving off the I_N domain (which should be visible in a SDS-PAGE) ultimately cyclizing the scFv making the cyclic scFv approximately 36 kDa. The linear version of the scFv does not contain any of the of the intein machinery and is approximately 31 kDa.

4.3.2.2 ScFv histidine tag-nickel column purification

A cyclic scFv was expressed as a fusion to 6x histidine tag using the pET28b expression plasmid. Cells expressing the cyclic scFv were induced and collected. The cells were ruptured by sonication. The cell lysate was denatured with 8 M urea and applied to the nickel beads. After an hour of incubation with the lysate, the beads were washed and eluted. Each sample was collected and separated in a SDS-PAGE (Figure 4-10).

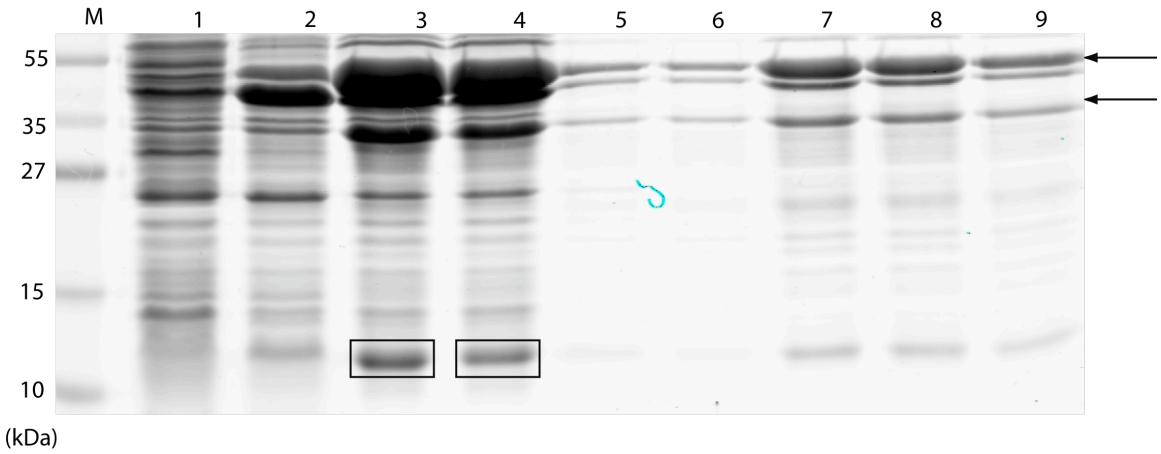


Figure 4-10: ScFv purification from *E. coli*.
Cells expressing a cyclic scFv were grown in LB-Kanamycin liquid media and induced with IPTG. A small sample of cells was taken before and after the induction and shown in the uninduced and induced lanes. The cell lysate was incubated with nickel beads. Lane M, standard pre-stained protein marker; lane 1, uninduced cells; lane 2, induced cells; lane 3, crude cell lysate before incubation with nickel beads; lane 4, flow through lysate that did not bind the nickel beads; lane 5, wash 1; lane 6, wash 2; lane 7, elution 1 from the nickel beads; lane 8, elution 8 from the nickel beads. Only the proteins containing a histidine tag will be visible in the elutions. The lower arrow indicates where the cyclic scFv is approximated to run (36 kDa). The upper arrow indicates where the unprocessed cyclic scFv is approximated to run (48 kDa). The boxes highlight the N-intein (12 kDa).

When the scFv was expressed in *E. coli* there are two main protein products purified from the nickel beads. There is a band that corresponds to the size of the cyclic scFv (36 kDa) and one that corresponds to the unprocessed scFv (48 kDa). The nickel beads were overloaded as indicated by the excess of protein in the flow through lane. There were bands in the crude and flow through samples that correspond to the size of the N-intein (12 kDa). The N-intein does not contain a histidine tag so it does not bind to the nickel beads as indicated by its presence only in the crude and flow through. The presence of the N-intein indicates that the intein is processing. Size exclusion was used to try and separate the three main bands from the histidine bead purification for further analysis of the cyclic scFv. The size exclusion was not successful in separating the bands. Further studies will have to be done to purify the cyclic scFv without the

contaminating bands in order to specifically analyze the cyclization. The three bands were analyzed on a mass spectrometer. The scFv was not visible on the mass spectrometer. The scFv may have been too large or not properly charged. Other groups who analyzed scFvs on the mass spectrometer have first digested the scFv prior to analysis (96). The proper conditions will have to be determined for analysis of the scFv on the mass spectrometer to determine if it is cyclic.

4.3.2.3 Renaturation of scFv

It has previously been shown that scFvs are difficult to purify in the native form from the cytoplasm of *E. coli* cells (97). To purify scFvs from the bacterial cytoplasm, most scFvs must first be denatured with urea or guanidium hydrochloride. When denaturing an scFv the ability to bind to protein L is eliminated. ScFvs will only bind to protein L when properly folded. To overcome this problem a basic renaturing protocol by Kou *et al*. was used to renature the scFvs (98). The scFv was denatured using 8 M urea and bound to the nickel beads (98). The scFv was renatured while still bound to the beads by washing it slowly with decreasing concentrations of urea. Most proteins cannot refold properly in a short amount of time (99). It must be slowly exposed to an environment in which it can start to refold (99).

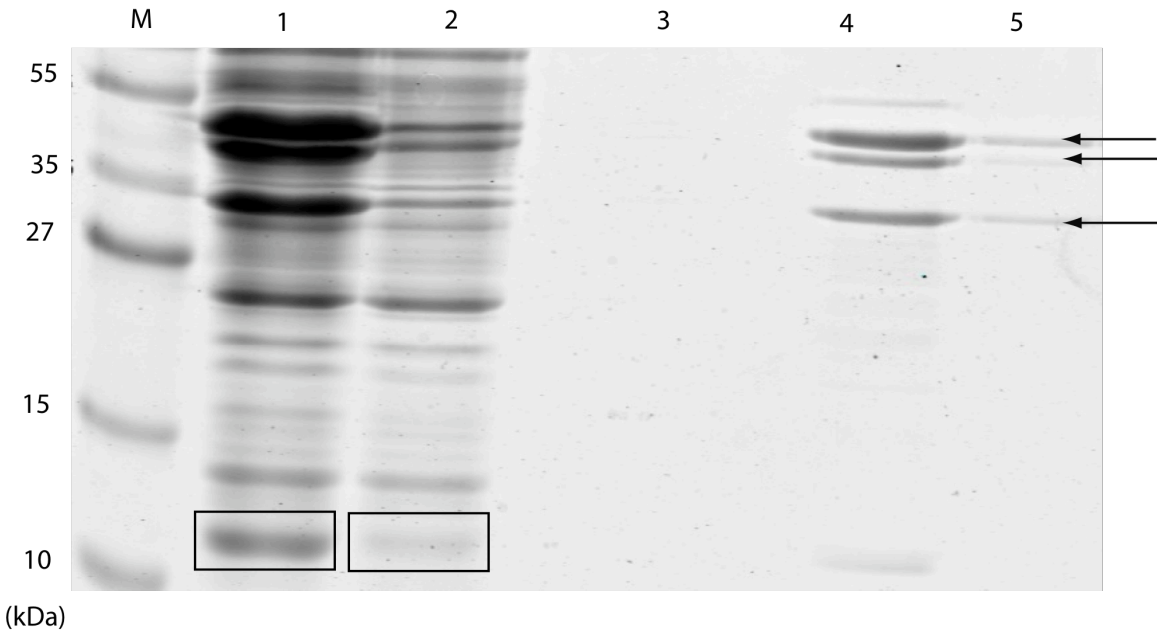


Figure 4-11: ScFv expression and renaturation.
A cyclic scFv was expressed in *E. coli*, grown in LB-Kanamycin liquid media and induced with IPTG. The cell lysate was incubated with nickel beads and washed with decreasing concentrations of urea in a 2 M gradient. Lane M, standard pre-stained protein marker; lane 1, crude cell lysate before incubation with nickel beads; lane 2, flow through lysate that did not bind the nickel beads; lane 3, wash 1; lane 4, wash 2; and lane 5, elution 1 from the nickel beads. Only the proteins containing a histidine tag will be visible in the elutions. The lower arrow indicates the cyclic scFv (36 kDa). The upper arrow indicates where the unprocessed cyclic scFv is (48 kDa). The boxes highlight the N-intein (12 kDa).

After washing the nickel beads with the scFv still bound, it was eluted. The samples from each step were separated in SDS-PAGE (Figure 4-11). There were three main bands eluted off the nickel beads after refolding. One corresponds to the cyclic scFv (36 kDa) and two correspond to the size of the unprocessed scFv (48 kDa). In the crude and flow through samples there is a band that corresponds to the size of the N-intein (12 kDa). It is impossible to tell from the SDS-PAGE what percent of the scFvs are renatured.

Previous studies have reported that <20% of the scFv will be refolded using this method. Further refolding occurred upon the addition of an oxidizing agent (100). ScFvs have been previously renatured; their activity has been restored upon refolding (98).

4.3.2.4 Protein L purification of scFv

Elution one from the renatured scFv was applied to the protein L beads and incubated overnight. The beads were washed, eluted and separated by SDS-PAGE (Figure 4-12).

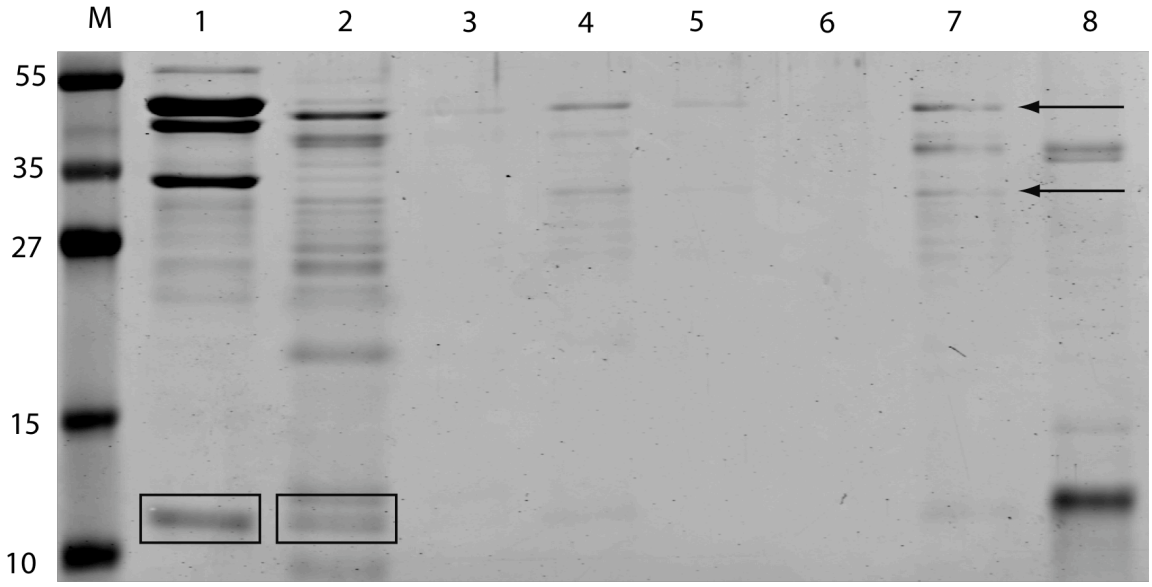


Figure 4-12: ScFv Protein L purification.
Elution one of renatured scFv from the nickel beads was incubated with protein L beads. After eluting the scFv off the protein L beads were heated up to 95 ˚C in loading dye to extract what was left on the beads. A control with only protein L beads was also heated in the same way. Lane M, standard pre-stained protein marker; lane 1, crude cell lysate before incubation with protein L beads; lane 2, flow through lysate that did not bind the protein L beads; lane 3 wash 1; lane 4, elution 1; lane 5 elution 2; lane 6, elution 3; lane 7, protein L beads heated after elution; lane 8, protein L beads heated alone. Only the proteins containing kappa light chains will bind to the protein L beads. The lower arrow indicates where the cyclic scFv is approximated to run (36 kDa). The upper arrow indicates where the unprocessed cyclic scFv is approximated to run (48 kDa). The boxes highlight the N-intein domain (12 kDa).

There was a low concentration of the scFv that binds to the protein L beads. There were two main bands that bind to the protein L column, which are visible in the loaded sample, the eluted sample and the protein L heated sample. There were other visible bands from the beads, if the bands are not visible in lane 1 they are artifacts from the protein L beads. The two main bands correspond to the cyclic scFv (36 kDa) and the unprocessed scFv (48 kDa). There was not very much protein eluting off the protein L beads. This could be due to a very small percentage of scFv actually refolding as has previously been shown (98). Refolding scFvs using only a decreasing urea gradient produces small yields of refolded scFv (>20%) (98). Using a decreasing gradient in correlation with an increasing pH gradient the refolded yield can be increased to 25% (101). By using a decreasing urea gradient in correlation with an oxidizing agent the yield of refolded scFv can be increased to as much as 75% if added at the right urea concentration (98).

Chapter 5 Discussion

By constructing large, diverse scFv libraries the effects of different parameters used to vary the CDR compositions and the scFv were studied to determine their impacts on binding capacity. In particular, cyclizing the scFv conformation allowed for more interactions with a set number of protein domains. This trend was reinforced by the statistically significant difference (P value < 0.05) between the cyclic and linear binding capacities.

The libraries with three variable CDRs had a higher binding capacity than the libraries with four CDRs in eleven out of thirteen screens. The increased diversity imposed by the fourth diversified CDR did not produce a larger binding capacity despite the increased diversity. The fourth diversified CDR may have created a negative effect on scFv folding and thus decreased the library binding capacity. In most of the screens the three CDR libraries had a statistically larger binding capacity than the four CDR libraries.

In a study by Fellouse *et al.,* two Fab phage display libraries were made, one with all three CDRs on the heavy chain varied and the other with the heavy chain CDRs and CDR L3 varied (9). The CDRs were diversified with tyrosine and serine. They found that most of the clones isolated from the library had the additional CDR L3 (9). The isolated clones from the same libraries had similar sequences but are different between the two libraries (9). This result concludes that the variable CDR L3 influences the specificity (9). In a similar study, four libraries were made; two were the same as described above and the other two were more diverse (81). The second least diverse library, the library with a variable heavy chain and variable L3

isolated the highest number of binders (81). These studies indicate that diversity may not dictate binding capacity.

The binding capacity results for the three CDR scFv libraries showed that the four amino acid libraries had a higher binding capacity then the two amino acid libraries. Three out of six screens had statistically significant differences (P value < 0.05). The binding capacity results for the four CDR libraries comparing two and four amino acid libraries were inconclusive. All sets of data showed a mix of results, which is signified by the random statistical significance of the results.

In a previous study, four different Fab libraries were created with the CDRs varied with tyrosine and serine (81). The simplest libraries had the heavy chain varied and the other had the heavy chain and light chain CDR L3 varied (81). The two more complex libraries contained additional variable amino acids throughout the CDRs (81). This is interesting because it does not reflect the most diverse library as obtaining the most interactions (81). These results indicate that a more diverse library in terms of CDR variability does not necessarily dictate more protein interactions.

Isolated scFvs were successfully expressed in *S. cerevisiae* and *E. coli*. The scFvs were not successfully isolated from *S. cerevisiae* with protein L. ScFvs have not previously been purified using protein L from *S. cerevisiae*. Many research groups purify scFvs from *S. cerevisiae* using metal chromatography and a histidine tag (102). Most groups use *Pichia pastoris* for scFv expression and purification instead of *S. cerevisiae*. They usually use a histidine tag to isolate the scFvs from the cell lysate (103,104) or after they are secreted into the media (105). A few groups have used protein L to isolate antibody fusions secreted into the media expressed in *Pichia pastoris* (106). It is unknown why the scFvs were not successfully

isolated using protein L. It is also interesting, since scFvs isolated with protein L from *S. cerevisiae* cells has not been previously reported. Previously Villani *et al*. reported purifying active scFvs from bacterial cytoplasm able to bind their target but unable to bind protein L for efficient purification (65).

ScFvs were successfully denatured and purified from the *E. coli* cytoplasm using nickel beads. ScFvs usually form insoluble inclusion bodies within *E. coli* cells and have to be denatured to purify them. To be functional scFvs must be properly refolded. There are very few reports on soluble scFvs isolated from the *E. coli* cytoplasm. They usually involve tethering of a soluble fusion protein (107); or more recently, using a soluble, stably expressed scFv framework (65). This scFv framework is able to bind its target *in vitro* after purification from the bacterial cytoplasm, however, is unable to bind to protein L (65). This indicates that protein L may not indicate proper folding of the scFvs since active scFvs were unable to bind to it. In most studies researchers use histidine tags to isolate scFvs from bacterial cell lysate or from the media the cells are growing in. When the scFvs are isolated from the cytoplasm they usually have to be denatured and then renatured to analyze activity or folding patterns. In many studies scFvs are secreted into the periplasm, which is an oxidizing environment that allows scFvs to fold properly. In the periplasm, expression level is often low and the formation of inclusion bodies is frequent despite its oxidizing character (108). The cytoplasm is a desirable expression compartment because although the scFvs are often unfolded, the concentration of proteins isolated is much higher (97).

The scFvs bound to the histidine beads were renatured and incubated with protein L beads. It was observed that very little protein bound to the protein L beads. Previous studies

report that <20 % of the scFv refold using this method (98). This may be the reason for such a small amount of protein eluted off the protein L column.

Chapter 6 Conclusion

In conclusion, cyclizing scFv libraries increases binding capacity. ScFv libraries with four amino acids produce a larger binding capacity than two amino acids in the three CDR libraries. Diversifying the fourth CDR, CDR L3 on the scFv decreases the binding capacity despite the increased diversity. The four CDR libraries had no trend when comparing the two and four amino acids. The goal of this project was to determine minimal parameters that could be varied in creating scFv libraries without sacrificing the ability to purify highly specific interactions. Minimal diversity limited to tyrosine and serine can be used with as little as 18 variable residues within the CDRs. With respect to library design if a new library were to be created, cyclization of the scFvs is recommended, which has been shown in this study to increase the binding ability. The new library would also be produced with three diversified CDRs and four amino acids. It would be interesting to determine the behavior of the specific scFvs within a human cell, where the target would be endogenous.

Chapter 7 Recommendations

Future recommendations for this project include optimizing the purification of scFvs to study the effect of cyclization. Purification in yeast can be performed using an HA-tag by denaturing the scFv before purification from yeast. Purification from *E. coli* involves optimizing the renaturing step, which can be achieved by adding an oxidizing reagent (98). Also, the three bands purified from cyclic scFvs expressed in *E. coli* will have to be separated to enable the study of the cyclic scFv alone. This can be achieved by cloning with two histidine tags, by using a more accurately size exclusion column, or by gel extraction.

This study has shown that the scFv is processing, which is a good indication that it is also cyclizing. This can be confirmed by studying the cyclic scFv with mass spectrometry. If the scFv is cyclic, the affect cyclization has on the scFv in terms of binding ability; affinity, stability and solubility must be determined. If the effect is as significant as shown in this study, problematic factors plaguing scFvs such as stability may be resolved. Binding assays can be done using surface plasmon resonance for example, to determine if the cyclic version is able to bind more effectively. Chemical and temperature unfolding and refolding studies can be performed to compare the stability of cyclic scFv and linear scFv. Solubility assays can be done to determine if cyclizing scFvs increases their solubility and reduces the formation of scFv multimers.

List of References

1. Reiter Y. & Pastan I. (1998). Recombinant Fv immunotoxins and Fv fragments as novel agents for cancer therapy and diagnosis. *Trends Biotechnol* **16**, 513-520.

2. Carlander D., Stålberg J. & Larsson A. (1999). Chicken antibodies: a clinical chemistry perspective. *Ups J Med Sci* **104**, 179-189.

3. Willats W. G. (2002). Phage display: practicalities and prospects. *Plant Mol Biol* **50**, 837-854.

4. Hoogenboom H. R. (2005). Selecting and screening recombinant antibody libraries. *Nat Biotechnol* **23**, 1105-1116.

5. Wörn A. & Plückthun A. (2001). Stability engineering of antibody single-chain Fv fragments. *J Mol Biol* **305**, 989-1010.

6. Pörtner-Taliana A., Russell M., Froning K. J., Budworth P. R., Comiskey J. D. & Hoeffler J. P. (2000). In vivo selection of single-chain antibodies using a yeast two-hybrid system. *J Immunol Methods* **238**, 161-172.

7. Coates P. J. & Hall P. A. (2003). The yeast two-hybrid system for identifying protein-protein interactions. *J Pathol* **199**, 4-7.

8. Fellouse F. A., Wiesmann C. & Sidhu S. S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc Natl Acad Sci U S A* **101**, 12467-12472.

9. Fellouse F. A., Li B., Compaan D. M., Peden A. A., Hymowitz S. G. & Sidhu S. S. (2005). Molecular recognition by a binary code. *J Mol Biol* **348**, 1153-1162.

10. Birtalan S., Zhang Y., Fellouse F. A., Shao L., Schaefer G. & Sidhu S. S. (2008). The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *J Mol Biol* **377**, 1518-1528.

11. Jurado P., de Lorenzo V. & Fernández L. A. (2006). Thioredoxin fusions increase folding of single chain Fv antibodies in the cytoplasm of Escherichia coli: evidence that chaperone activity is the prime effect of thioredoxin. *J Mol Biol* **357**, 49-61.

12. Iwai H. & Plückthun A. (1999). Circular beta-lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett* **459**, 166-172.

13. Iwai H., Lingel A. & Pluckthun A. (2001). Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfuI intein from Pyrococcus furiosus. *J Biol Chem* **276**, 16548-16554.

14. Camarero J. A., Fushman D., Sato S., Giriat I., Cowburn D., Raleigh D. P. & Muir T. W. (2001). Rescuing a destabilized protein fold through backbone cyclization. *J Mol Biol* **308**, 1045-1062.

15. Williams N. K., Liepinsh E., Watt S. J., Prosselkov P., Matthews J. M., Attard P., Beck J. L., Dixon N. E. & Otting G. (2005). Stabilization of native protein fold by intein-mediated covalent cyclization. *J Mol Biol* **346**, 1095-1108.

16. Lobato M. N. & Rabbitts T. H. (2004). Intracellular antibodies as specific reagents for functional ablation: future therapeutic molecules. *Curr Mol Med* **4**, 519-528.

17. Antman K. H. & Livingston D. M. (1980). Intracellular neutralization of SV40 tumor antigens following microinjection of specific antibody. *Cell* **19**, 627-635.

18. Wirtz P. & Steipe B. (1999). Intrabody construction and expression III: engineering hyperstable V(H) domains. *Protein Sci* **8**, 2245-2250.

19. Skerra A. & Plückthun A. (1988). Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science* **240**, 1038-1041.

20. Freund C., Ross A., Guth B., Plückthun A. & Holak T. A. (1993). Characterization of the linker peptide of the single-chain Fv fragment of an antibody by NMR spectroscopy. *FEBS Lett* **320**, 97-100.

21. Jääskeläinen A., Harinen R. R., Soukka T., Lamminmäki U., Korpimäki T. & Virta M. (2008). Biologically produced bifunctional recombinant protein nanoparticles for immunoassays. *Anal Chem* **80**, 583-587.

22. Carlsson A., Wingren C., Ingvarsson J., Ellmark P., Baldertorp B., Fernö M., Olsson H. & Borrebaeck C. A. (2008). Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. *Eur J Cancer* **44**, 472-480.

23. Steinhauer C., Wingren C., Hager A. C. & Borrebaeck C. A. (2002). Single framework recombinant antibody fragments designed for protein chip applications. *Biotechniques* **Suppl**, 38-45.

24. Mousli M., Turki I., Kharmachi H., Saadi M. & Dellagi K. (2007). Recombinant single-chain Fv antibody fragment-alkaline phosphatase conjugate: a novel in vitro tool to estimate rabies viral glycoprotein antigen in vaccine manufacture. *J Virol Methods* **146**, 246-256.

25. Rincón V., Corredor A., Martínez-Gutíerrez M. & Castellanos J. E. (2005). Fluorometric cell-ELISA for quantifying rabies infection and heparin inhibition. *J Virol Methods* **127**, 33-39.

26. Laroche Y., Demaeyer M., Stassen J. M., Gansemans Y., Demarsin E., Matthyssens G., Collen D. & Holvoet P. (1991). Characterization of a recombinant single-chain molecule comprising the variable domains of a monoclonal antibody specific for human fibrin fragment D-dimer. *J Biol Chem* **266**, 16343-16349.

27. Samuel A. S. & Naz R. K. (2008). Isolation of human single chain variable fragment antibodies against specific sperm antigens for immunocontraceptive development. *Hum Reprod* **23**, 1324-1337.

28. Wels W., Harwerth I. M., Mueller M., Groner B. & Hynes N. E. (1992). Selective inhibition of tumor cell growth by a recombinant single-chain antibody-toxin specific for the erbB-2 receptor. *Cancer Res* **52**, 6310-6317.

29. www.herceptin.com

30. Mian I. S., Bradwell A. R. & Olson A. J. (1991). Structure, function and properties of antibody binding sites. *J Mol Biol* **217**, 133-151.

31. Davies D. R. & Cohen G. H. (1996). Interactions of protein antigens with antibodies. *Proc Natl Acad Sci U S A* **93**, 7-12.

32. Sidhu S. S., Li B., Chen Y., Fellouse F. A., Eigenbrot C. & Fuh G. (2004). Phage-displayed antibody libraries of synthetic heavy chain complementarity determining regions. *J Mol Biol* **338**, 299-310.

33. Zemlin M., Klinger M., Link J., Zemlin C., Bauer K., Engler J. A., Schroeder H. W. & Kirkham P. M. (2003). Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. *J Mol Biol* **334**, 733-749.

34. Nikula T. K., Bocchia M., Curcio M. J., Sgouros G., Ma Y., Finn R. D. & Scheinberg D. A. (1995). Impact of the high tyrosine fraction in complementarity determining regions: measured and predicted effects of radioiodination on IgG immunoreactivity. *Mol Immunol* **32**, 865-872.

35. Collis A. V., Brouwer A. P. & Martin A. C. (2003). Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. *J Mol Biol* **325**, 337-354.

36. Katz M. & Papenheimer A. M. (1969). Quantitative studies of the specificity of anti-pneumococcal antibodies, types 3 and 8. IV. Binding of labeled hexasaccharides derived from S3 by anti-S3 antibodies and their Fab fragments. *J Immunol* **103**, 491-495.

37. Proba K., Honegger A. & Plückthun A. (1997). A natural antibody missing a cysteine in VH: consequences for thermodynamic stability and folding. *J Mol Biol* **265**, 161-172.

38. Röthlisberger D., Honegger A. & Plückthun A. (2005). Domain interactions in the Fab fragment: a comparative evaluation of the single-chain Fv and Fab format engineered with variable domains of different stability. *J Mol Biol* **347**, 773-789.

39. Desplancq D., King D. J., Lawson A. D. & Mountain A. (1994). Multimerization behaviour of single chain Fv variants for the tumour-binding antibody B72.3. *Protein Eng* **7**, 1027-1033.

40. Whitlow M., Filpula D., Rollence M. L., Feng S. L. & Wood J. F. (1994). Multivalent Fvs: characterization of single-chain Fv oligomers and preparation of a bispecific Fv. *Protein Eng* **7**, 1017-1026.

41. Reiter Y., Brinkmann U., Kreitman R. J., Jung S. H., Lee B. & Pastan I. (1994). Stabilization of the Fv fragments in recombinant immunotoxins by disulfide bonds engineered into conserved framework regions. *Biochemistry* **33**, 5451-5459.

42. Robinson C. R. & Sauer R. T. (1998). Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proc Natl Acad Sci U S A* **95**, 5929-5934.

43. Shaki-Loewenstein S., Zfania R., Hyland S., Wels W. S. & Benhar I. (2005). A universal strategy for stable intracellular antibodies. *J Immunol Methods* **303**, 19-39.

44. Nieba L., Honegger A., Krebber C. & Plückthun A. (1997). Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved in vivo folding and physical characterization of an engineered scFv fragment. *Protein Eng* **10**, 435-444.

45. Martineau P., Jones P. & Winter G. (1998). Expression of an antibody fragment at high levels in the bacterial cytoplasm. *J Mol Biol* **280**, 117-127.

46. Philibert P., Stoessel A., Wang W., Sibler A. P., Bec N., Larroque C., Saven J. G., Courtête J., Weiss E. & Martineau P. (2007). A focused antibody library for selecting scFvs expressed at high levels in the cytoplasm. *BMC Biotechnol* **7**, 81.

47. Kim Y. J., Neelamegam R., Heo M. A., Edwardraja S., Paik H. J. & Lee S. G. (2008). Improving the Productivity of Single-Chain Fv Antibody Against c-Met by Rearranging the Order of its Variable Domains. *J Microbiol Biotechnol* **18**, 1186-1190.

48. Holliger P., Prospero T. & Winter G. (1993). "Diabodies": small bivalent and bispecific antibody fragments. *Proc Natl Acad Sci U S A* **90**, 6444-6448.

49. Visintin M., Tse E., Axelson H., Rabbitts T. H. & Cattaneo A. (1999). Selection of antibodies for intracellular function using a two-hybrid in vivo system. *Proc Natl Acad Sci U S A* **96**, 11723-11728.

50. Jung S., Honegger A. & Plückthun A. (1999). Selection for improved protein stability by phage display. *J Mol Biol* **294**, 163-180.

51. Björck L. (1988). Protein L. A novel bacterial cell wall protein with affinity for Ig L chains. *J Immunol* **140**, 1194-1197.

52. Akerström B. & Björck L. (1989). Protein L: an immunoglobulin light chain-binding bacterial protein. Characterization of binding and physicochemical properties. *J Biol Chem* **264**, 19740-19746.

53. Akerström B., Nilson B. H., Hoogenboom H. R. & Björck L. (1994). On the interaction between single chain Fv antibodies and bacterial immunoglobulin-binding proteins. *J Immunol Methods* **177**, 151-163.

54. Scott C. P., Abel-Santos E., Wall M., Wahnon D. C. & Benkovic S. J. (1999). Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci U S A* **96**, 13638-13643.

55. Evans T. C., Martin D., Kolly R., Panne D., Sun L., Ghosh I., Chen L., Benner J., Liu X. Q. & Xu M. Q. (2000). Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of Synechocystis species PCC6803. *J Biol Chem* **275**, 9091-9094.

56. Liu X. Q. (2000). Protein-splicing intein: Genetic mobility, origin, and evolution. *Annu Rev Genet* **34**, 61-76.

57. Wu H., Hu Z. & Liu X. Q. (1998). Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc Natl Acad Sci U S A* **95**, 9226-9231.

58. Iwai H., Züger S., Jin J. & Tam P. H. (2006). Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Lett* **580**, 1853-1858.

59. Jeffries C. M., Graham S. C., Stokes P. H., Collyer C. A., Guss J. M. & Matthews J. M. (2006). Stabilization of a binary protein complex by intein-mediated cyclization. *Protein Sci* **15**, 2612-2618.

60. Huston J. S., Levinson D., Mudgett-Hunter M., Tai M. S., Novotný J., Margolies M. N., Ridge R. J., Bruccoleri R. E., Haber E. & Crea R. (1988). Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. *Proc Natl Acad Sci U S A* **85**, 5879-5883.

61. Bird R. E., Hardman K. D., Jacobson J. W., Johnson S., Kaufman B. M., Lee S. M., Lee T., Pope S. H., Riordan G. S. & Whitlow M. (1988). Single-chain antigen-binding proteins. *Science* **242**, 423-426.

62. Wörn A., Auf der Maur A., Escher D., Honegger A., Barberis A. & Plückthun A. (2000). Correlation between in vitro stability and in vivo performance of anti-GCN4 intrabodies as cytoplasmic inhibitors. *J Biol Chem* **275**, 2795-2803.

63. De Jaeger G., Fiers E., Eeckhout D. & Depicker A. (2000). Analysis of the interaction between single-chain variable fragments and their antigen in a reducing intracellular environment using the two-hybrid system. *FEBS Lett* **467**, 316-320.

64. Gennari F., Mehta S., Wang Y., St Clair Tallarico A., Palu G. & Marasco W. A. (2004). Direct phage to intrabody screening (DPIS): demonstration by isolation of cytosolic intrabodies against the TES1 site of Epstein Barr virus latent membrane protein 1 (LMP1) that block NF-kappaB transactivation. *J Mol Biol* **335**, 193-207.

65. Villani M. E., Di Carli M., Donini M., Traversini G., Lico C., Franconi R., Benvenuto E. & Desiderio A. (2008). Validation of a stable recombinant antibodies repertoire for the direct selection of functional intracellular reagents. *J Immunol Methods* **329**, 11-20.

66. Fields S. & Song O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246.

67. Golemis E. A. & Brent R. (1992). Fused protein domains inhibit DNA binding by LexA. *Mol Cell Biol* **12**, 3006-3014.

68. Cohen B. A., Colas P. & Brent R. (1998). An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* **95**, 14272-14277.

69. Geyer C. R. & Brent R. (2000). Selection of genetic agents from random peptide aptamer expression libraries. *Methods Enzymol* **328**, 171-208.

70. Elbing K. & Brent R. (2002). Media preparation and bacteriological tools. *Curr Protoc Mol Biol* **Chapter 1**, Unit 1.1.

71. Elbing K. & Brent R. (2002). Growth in liquid media. *Curr Protoc Mol Biol* **Chapter 1**, Unit 1.2.

72. Elbing K. & Brent R. (2002). Growth on solid media. *Curr Protoc Mol Biol* **Chapter 1**, Unit 1.3.

73. Tanaka T., Chung G. T., Forster A., Lobato M. N. & Rabbitts T. H. (2003). De novo production of diverse intracellular antibody libraries. *Nucleic Acids Res* **31**, e23.

74. Tse E., Lobato M. N., Forster A., Tanaka T., Chung G. T. & Rabbitts T. H. (2002). Intracellular antibody capture technology: application to selection of intracellular antibodies recognising the BCR-ABL oncogenic protein. *J Mol Biol* **317**, 85-94.

75. http://slam.bs.jhmi.edu/gd/

76. Schiestl R. H. & Gietz R. D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet* **16**, 339-346.

77. Kabat E. A. & Wu T. T. (1991). Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. *J Immunol* **147**, 1709-1719.

78. Pluthero F. G. (1993). Rapid purification of high-activity Taq DNA polymerase. *Nucleic Acids Res* **21**, 4850-4851.

79. Kushnirov V. V. (2000). Rapid and reliable protein extraction from yeast. *Yeast* **16**, 857-860.

80. Dower W. J., Miller J. F. & Ragsdale C. W. (1988). High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res* **16**, 6127-6145.

81. Fellouse F. A., Esaki K., Birtalan S., Raptis D., Cancasci V. J., Koide A., Jhurani P., Vasser M., Wiesmann C., Kossiakoff A. A., Koide S. & Sidhu S. S. (2007). High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J Mol Biol* **373**, 924-940.

82. Davies D. R., Padlan E. A. & Sheriff S. (1990). Antibody-antigen complexes. *Annu Rev Biochem* **59**, 439-473.

83. Bickle M. B., Dusserre E., Moncorgé O., Bottin H. & Colas P. (2006). Selection and characterization of large collections of peptide aptamers through optimized yeast two-hybrid procedures. *Nat Protoc* **1**, 1066-1091.

84. Heisterkamp N., Stam K., Groffen J., de Klein A. & Grosveld G. (1985). Structural organization of the bcr gene and its role in the Ph' translocation. *Nature* **315**, 758-761.

85. Danhauser-Riedl S., Warmuth M., Druker B. J., Emmerich B. & Hallek M. (1996). Activation of Src kinases p53/56lyn and p59hck by p210bcr/abl in myeloid cells. *Cancer Res* **56**, 3589-3596.

86. Stanglmaier M., Warmuth M., Kleinlein I., Reis S. & Hallek M. (2003). The interaction of the Bcr-Abl tyrosine kinase with the Src kinase Hck is mediated by multiple binding domains. *Leukemia* **17**, 283-289.

87. Zhao X., Ghaffari S., Lodish H., Malashkevich V. N. & Kim P. S. (2002). Structure of the Bcr-Abl oncoprotein oligomerization domain. *Nat Struct Biol* **9**, 117-120.

88. Overduin M., Rios C. B., Mayer B. J., Baltimore D. & Cowburn D. (1992). Three-dimensional solution structure of the src homology 2 domain of c-abl. *Cell* **70**, 697-704.

89. Nam H. J., Haser W. G., Roberts T. M. & Frederick C. A. (1996). Intramolecular interactions of the regulatory domains of the Bcr-Abl kinase reveal a novel control mechanism. *Structure* **4**, 1105-1114.

90. Chu S., Li L., Singh H. & Bhatia R. (2007). BCR-tyrosine 177 plays an essential role in Ras and Akt activation and in human hematopoietic progenitor transformation in chronic myelogenous leukemia. *Cancer Res* **67**, 7045-7053.

91. Ma G., Lu D., Wu Y., Liu J. & Arlinghaus R. B. (1997). Bcr phosphorylated on tyrosine 177 binds Grb2. *Oncogene* **14**, 2367-2372.

92. Heaney C., Kolibaba K., Bhat A., Oda T., Ohno S., Fanning S. & Druker B. J. (1997). Direct binding of CRKL to BCR-ABL is not required for BCR-ABL transformation. *Blood* **89**, 297-306.

93. Hantschel O., Wiesner S., Güttler T., Mackereth C. D., Rix L. L., Mikes Z., Dehne J., Görlich D., Sattler M. & Superti-Furga G. (2005). Structural basis for the cytoskeletal association of Bcr-Abl/c-Abl. *Mol Cell* **19**, 461-473.

94. Miller K. D., Weaver-Feldhaus J., Gray S. A., Siegel R. W. & Feldhaus M. J. (2005). Production, purification, and characterization of human scFv antibodies expressed in Saccharomyces cerevisiae, Pichia pastoris, and Escherichia coli. *Protein Expr Purif* **42**, 255-267.

95. Robin S., Petrov K., Dintinger T., Kujumdzieva A., Tellier C. & Dion M. (2003). Comparison of three microbial hosts for the expression of an active catalytic scFv. *Mol Immunol* **39**, 729-738.

96. Das D., Allen T. M. & Suresh M. R. (2005). Comparative evaluation of two purification methods of anti-CD19-c-myc-His6-Cys scFv. *Protein Expr Purif* **39**, 199-208.

97. Dueñas M., Vázquez J., Ayala M., Söderlind E., Ohlin M., Pérez L., Borrebaeck C. A. & Gavilondo J. V. (1994). Intra- and extracellular expression of an scFv antibody fragment in E. coli: effect of bacterial strains and pathway engineering using GroES/L chaperonins. *Biotechniques* **16**, 476-7, 480-3.

98. Kou G., Shi S., Wang H., Tan M., Xue J., Zhang D., Hou S., Qian W., Wang S., Dai J., Li B. & Guo Y. (2007). Preparation and characterization of recombinant protein ScFv(CD11c)-TRP2 for tumor therapy from inclusion bodies in Escherichia coli. *Protein Expr Purif* **52**, 131-138.

99. Teipel J. W. (1972). In vitro assembly of aldolase. Kinetics of refolding, subunit reassociation, and reactivation. *Biochemistry* **11**, 4100-4107.

100. Guo J. Q., You S. Y., Li L., Zhang Y. Z., Huang J. N. & Zhang C. Y. (2003). Construction and high-level expression of a single-chain Fv antibody fragment specific for acidic isoferritin in Escherichia coli. *J Biotechnol* **102**, 177-189.

101. Yang X., Hu W., Li F., Xia H. & Zhang Z. (2005). Gene cloning, bacterial expression, in vitro refolding, and characterization of a single-chain Fv antibody against PreS1(21-47) fragment of HBsAg. *Protein Expr Purif* **41**, 341-348.

102. Hackel B. J., Huang D., Bubolz J. C., Wang X. X. & Shusta E. V. (2006). Production of soluble and active transferrin receptor-targeting single-chain antibody using Saccharomyces cerevisiae. *Pharm Res* **23**, 790-797.

103. Peterson E. C., Laurenzana E. M., Atchley W. T., Hendrickson H. P. & Owens S. M. (2008). Development and preclinical testing of a high-affinity single-chain antibody against (+)-methamphetamine. *J Pharmacol Exp Ther* **325**, 124-133.

104. Ettayebi K. & Hardy M. E. (2008). Recombinant norovirus-specific scFv inhibit virus-like particle binding to cellular ligands. *Virol J* **5**, 21.

105. Ren F., Li B. C., Zhang N. N., Cao M., Dan W. B. & Zhang S. Q. (2008). Expression, purification and characterization of anti-BAFF antibody secreted from the yeast Pichia pastoris. *Biotechnol Lett* **30**, 1075-1080.

106. Han X., Ye L. B., Li B. Z., Bo G., Cai W. J., Hong Z., She Y. L., Li Y., Kong L. B. & Wu Z. H. (2006). Expression, purification and characterization of the Hepatitis B virus entire envelope large protein in Pichia pastoris. *Protein Expr Purif* **49**, 168-175.

107. Bach H., Mazor Y., Shaky S., Shoham-Lev A., Berdichevsky Y., Gutnick D. L. & Benhar I. (2001). Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *J Mol Biol* **312**, 79-93.

108. Kipriyanov S. M., Dübel S., Breitling F., Kontermann R. E., Heymann S. & Little M. (1995). Bacterial expression and refolding of single-chain Fv fragments with C-terminal cysteines. *Cell Biophys* **26**, 187-204.

Appendix

Raw Screen Data

Table A-1 Raw Screen Data.
This is the raw data from each screen, the numbers counted from each plate. They have not been normalized to functionality or actual diploids plated, which was counted twice in triplicate. The first column shows the library name linear =L or cyclic = c, K = four amino acids or T = two amino acids. The numbers from screens 1, 2, 3, 4, 5 are shown in the few columns. The average of the screens is shown in the seventh column. The standard deviation (SD) is shown in the eighth column. The number of diploids screened is shown in the next column. The functional diversity as defined by correct sequence is shown in the second last column. The interactions per functional diversity (FD) multiplied by ten million is shown in the last column.

Screen 1	1	2	3	4	5	Average	SD	Diploids	Functional Diversity	Interactions/ FD*1x10^7
LK3	882	902	904	847	842	875	30	2.60E+05	0.73	4612.2
cK3	NA	NA	NA	NA	NA	NA	NA	NA	0.60	NA
LK4	404	124	105	131	101	173	130	2.60E+05	0.52	1279.6
cK4	304	403	402	403	262	355	67	2.60E+05	0.41	3328.3
LT3	108	58	84	98	75	85	20	2.60E+05	0.67	485.6
cT3	NA	NA	NA	NA	NA	NA	NA	NA	0.60	NA
LT4	6	21	32	7	11	15	11	2.60E+05	0.51	116.1
cT4	264	185	102	410	358	264	125	2.60E+05	0.65	1560.9

Screen 2	1	2	3	4	5	Average	SD	Diploids	Functional Diversity	Interactions/ FD*1x10^7
LK3	296	112	0	0	ND	102	140	1.00E+06	0.73	139.7
cK3	NA	NA	NA	NA	NA	NA	NA	NA	0.60	NA
LK4	168	ND	92	50	42	88	58	1.00E+06	0.52	169.2
cK4	148	156	50	148	74	115	49	1.00E+06	0.41	281.0
LT3	26	52	46	35	19	36	14	1.00E+06	0.67	53.1
cT3	NA	NA	NA	NA	NA	NA	NA	NA	0.60	NA
LT4	92	59	47	30	65	59	23	1.00E+06	0.51	114.9
cT4	184	144	168	128	ND	156	25	1.00E+06	0.65	240.0

Screen 3	1	2	3	4	5	Average	SD	Diploids	Functional Diversity	Interactions/ FD*1x10^7
LK3	84	97	125	193	138	127	43	6.00E+06	0.73	29.1
cK3	NA	NA	NA	NA	NA	NA	NA	NA	0.60	NA
LK4	49	54	67	51	64	57	8	6.00E+06	0.52	18.3
cK4	53	55	68	69	80	65	11	6.00E+06	0.41	26.4
LT3	119	124	160	170	147	144	22	6.00E+06	0.67	35.8
cT3	NA	NA	NA	NA	NA	NA	NA	NA	0.60	NA
LT4	80	88	118	94	87	93	15	6.00E+06	0.51	30.5
cT4	147	158	163	154	201	165	21	6.00E+06	0.65	42.2

Screen 4	1	2	3	4	5	Average	SD	Diploids	Functional Diversity	Interactions/ FD*1x10^7
LK3	104	100	129	95	107	107	13	2.02E+06	0.73	72.6
cK3	NA	NA	NA	NA	NA	NA	NA	NA	0.60	NA
LK4	23	15	24	26	22	22	4	2.37E+06	0.52	17.9
cK4	83	112	106	89	ND	98	14	1.90E+06	0.41	125.2
LT3	41	40	75	21	37	43	20	1.93E+06	0.67	33.1
cT3	73	42	42	88	61	61	20	1.75E+06	0.60	58.3
LT4	31	20	17	14	ND	21	7	1.33E+06	0.51	30.2
cT4	54	82	36	54	57	57	16	2.13E+06	0.65	40.9

Screen 5	1	2	3	4	5	Average	SD	Diploids	Functional Diversity	Interactions/ FD*1x10^7
LK3	84	163	130	123	107	121	29	3.32E+07	0.73	50.1
cK3	263	211	231	190	216	222	27	2.92E+07	0.60	126.8
LK4	19	17	13	9	27	17	7	4.59E+07	0.52	7.1
cK4	47	52	30	31	33	39	10	4.63E+07	0.41	20.3
LT3	20	23	17	19	27	21	4	3.31E+07	0.67	9.6
cT3	31	39	36	33	45	37	5	3.77E+07	0.60	16.3
LT4	13	7	19	8	9	11	5	3.67E+07	0.51	6.0
cT4	44	43	52	47	54	48	5	3.14E+07	0.65	23.5